

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: March 17, 2003, 12:13:06 ; Search time 10.3726 Seconds
(without alignments)
3378.605 Million cell updates/sec

Title: US-09-964-858-1_COPY_1_263
Perfect score: 1386
Sequence: 1 MNSYPSKLPIDKSHSLQLQ.....NKNNEVNSEPALTDMLKR 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
A_Geneseq_101002:*

1:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1386	100.0	1664	20	AAW99462
2	1386	100.0	1664	23	AAU79331
3	1386	100.0	1664	23	AAE19799
4	241	17.4	236	20	AAW99456
5	241	17.4	236	23	AAE19800
6	138.5	10.0	1166	20	AAV08643
7	133	9.6	767	22	ABBS8240
8	133	9.6	971	19	AAW48896
9	131.5	9.5	1174	22	AAG85039
10	131.5	9.5	2439	22	ABB68896

11	127.5	9.2	1335	23	ABP38871
12	127.5	9.2	1542	22	ABB71456
13	127	9.2	4134	20	AAV31946
14	126	9.1	1817	21	AAW42230
15	125	9.0	722	22	ABB63899
16	125	9.0	1545	22	ABB65577
17	124.5	9.0	811	22	ABB62661
18	124	8.9	1798	22	ABB71695
19	124	8.9	2781	21	AAV57453
20	123	8.9	672	22	AAW93543
21	123	8.9	874	22	ABB62601
22	123	8.9	1557	22	ABB63735
23	123	8.9	4019	22	AAE13839
24	122.5	8.8	2703	22	ABB60074
25	122	8.8	761	22	ABB67519
26	120.5	8.7	1069	22	ABB61305
27	120.5	8.7	1428	22	ABB70377
28	120	8.7	844	22	AAU35904
29	120	8.7	849	18	AAW20655
30	120	8.7	1744	22	ABB62890
31	120	8.7	1908	22	ABB70137
32	120	8.7	1954	22	ABB59197
33	119.5	8.6	690	22	ABB71683
34	119.5	8.6	1888	22	AAU31049
35	119.5	8.6	1920	22	ABB64441
36	119	8.6	1186	22	ABB64079
37	118.5	8.5	696	22	ABB60920
38	118	8.5	738	22	ABB58060
39	118	8.5	2907	21	AAV57452
40	118	8.5	10182	23	ABP38314
41	117.5	8.5	3502	22	ABB58382
42	117	8.4	5464	22	ABG11810
43	117	8.4	7718	22	ABG11811
44	116.5	8.4	1379	22	ABB68840
45	116	8.4	797	22	ABB62042

ALIGNMENTS

RESULT 1
AAW99462
AAW99462 standard; Protein: 1664 AA.

AAW99462:
08-JUN-1999 (first entry)

C.albicans alpha-INTip protein.
Integrin-like motif; vaccine; immune response; antibody; inhibition;
adhesion; endothelial cell; pathogenesis; infection; probe.

Candida albicans.
US5886151-A.
23-MAR-1999.
03-MAY-1996; 96US-0642846.
03-MAY-1996; 96US-0642846.
03-MAY-1996; 96US-0642846.
(MIND) UNIV MINNESOTA.
Bendel CM, Gale CA, Hostetter MK, Kendrick K, Tao NJ;
WPI; 1999-242618/20.
N-PSDB; AA25885.
New isolated Candida albicans protein with Integrin-like motifs
Examples: Column 13-14; 21pp; English.

Staphylococcus epi
Drosophila melanog
Plasmodium falcipa
Human ORF1994
Drosophila melanog
Drosophila melanog
Drosophila melanog
Drosophila melanog
Human transcriptio
Human polypeptide,
Drosophila melanog
Drosophila melanog
Human lung tumour-
Drosophila melanog
Drosophila melanog
Drosophila melanog
Helicobacter pylori
H. pylori cytoplas
Drosophila melanog
Drosophila melanog
Drosophila melanog
Drosophila melanog
Novel human secre
Drosophila melanog
Drosophila melanog
Drosophila melanog
Human transcriptio
Staphylococcus epi
Drosophila melanog
Novel human diagno
Drosophila melanog
Drosophila melanog

XX This sequence represents the *Candida albicans* alpha-INT1 protein which
 CC contains integrin-like motifs. The protein was used to derive peptides
 CC AA099456-W99461 used for producing vaccines for stimulating an immune
 CC response. The antibodies can inhibit the adhesion of *C. albicans* to
 CC cells, particularly endothelial cells. This blocking activity of the
 CC adhesion to cells can reduce or prevent subsequent events in the
 CC pathogenesis of invasive candidal infection.

XX Sequence 1664 AA:

Query Match 100.0%; Score 1386; DB 20; Length 1664;
 Best Local Similarity 100.0%; Pred. No. 5e-107;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSTPSKLLPDKHSHLQLOPOSSASIFNSPTKPLNPRNRSKPSLDPNSSDYTSQ 60
 DB 1 MNSTPSKLLPDKHSHLQLOPOSSASIFNSPTKPLNPRNRSKPSLDPNSSDYTSQ 60
 61 DOEKGEKKKDTAFQTSFDRNFDLNSIDIOQTIGHQOQPOQOQOLSTQDNLLIDFSF 120
 DB 61 DOEKGEKKKDTAFQTSFDRNFDLNSIDIOQTIGHQOQPOQOQOLSTQDNLLIDFSF 120
 QY 121 QTPMTSTDLTKQNTYVKVNHAPTYINTSPKSIKKATPKKSPKVAFTVTPNPEH 180
 DB 121 QTPMTSTDLTKQNTYVKVNHAPTYINTSPKSIKKATPKKSPKVAFTVTPNPEH 180
 QY 181 HYPDNRVEEEDSOQKEDSVPEPLIOHOKDPSQFNYSDEDTNASVPTPLHTTKPTFA 240
 DB 181 HYPDNRVEEEDSOQKEDSVPEPLIOHOKDPSQFNYSDEDTNASVPTPLHTTKPTFA 240
 QY 241 QLNKNNEVNSEPALTDMLKR 263
 DB 241 QLNKNNEVNSEPALTDMLKR 263

RESULT 2

AA079331

ID AA079331 standard; Protein: 1664 AA.

XX AA079331;

XX 02-JUN-2002 (first entry)

XX *Candida albicans* integrin-like protein, Intlp propeptide.
 DE Integrin-like protein; Intlp propeptide; bactericide; vaccine;
 XX yeast infection; thrush; paronychia; candidal vaginitis;
 XX immunocompromised host; cancer patient; transplant patient;
 XX premature newborn; human immunodeficiency virus infection;
 XX HIV infection; periodontitis; oral ulceration; esophagitis.

XX *Candida albicans*.

XX Key Location/Qualifiers

FT Binding-site 155..169

FT /label=Heparin_binding_site

FT Domain 435..639

FT /label=Catalytic_domain_1

FT Domain 738..949

FT /label=Catalytic_domain_2

FT Domain 1022..1236

FT /label=P-domain

FT /note="Processing domain"

XX W0200226257-A1.

XX 04-APR-2002.

XX 28-SEP-2001; 2001WO-US30312.

XX 28-SEP-2000; 2000US-237082P.

PA (UYVA) UNIV YALE.

XX Hostetter MK, Devore-Carter D;

XX WPI: 2002-330056/36.

DR N-PSDB; ABK48292.

PT Antibodies that react with the Intlp pro-peptide or peptide, useful for
 PT vaccinating against infections by *Candida albicans* and *Saccharomyces*
 PT *cerevisiae*.

PS Disclosure: Fig 2A-B; 53pp; English.

CC The invention describes an isolated antibody generated against the
 CC pro-peptide of the Intlp protein of *Candida albicans* or a peptide region
 CC of the Intlp protein of *C. albicans* involved in the activation of the
 CC Intlp pro-peptide. The antibody may be administered for the treatment of
 CC diseases caused by *C. albicans* and *Saccharomyces cerevisiae* (e.g. in the
 CC form of a vaccine) and for detecting the presence of Intlp polypeptides
 CC in samples for the diagnosis of these diseases. The antibody can be used
 CC to treat *C. albicans* infections including thrush in newborn infants,
 CC paronychia and candidal vaginitis. The antibody can also be used to treat
 CC *C. albicans* infections in immunocompromised hosts such as cancer
 CC patients, transplant patients, premature newborns or patients infected
 CC with human immunodeficiency virus (HIV) on which *C. albicans* causes
 CC diseases including periodontitis, oral ulceration and esophagitis. This
 CC is the amino acid sequence of the *C. albicans* integrin-like protein,
 CC Intlp propeptide that plays a major role in activating T-lymphocytes and
 CC can be cleaved to release a superantigen-like moiety.

SQ Sequence 1664 AA:

Query Match 100.0%; Score 1386; DB 23; Length 1664;
 Best Local Similarity 100.0%; Pred. No. 5e-107;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSTPSKLLPDKHSHLQLOPOSSASIFNSPTKPLNPRNRSKPSLDPNSSDYTSQ 60
 DB 1 MNSTPSKLLPDKHSHLQLOPOSSASIFNSPTKPLNPRNRSKPSLDPNSSDYTSQ 60
 QY 61 DOEKGEKKKDTAFQTSFDRNFDLNSIDIOQTIGHQOQPOQOQOLSTQDNLLIDFSF 120
 DB 61 DOEKGEKKKDTAFQTSFDRNFDLNSIDIOQTIGHQOQPOQOQOLSTQDNLLIDFSF 120
 QY 121 QTPMTSTDLTKQNTYVKVNHAPTYINTSPKSIKKATPKKSPKVAFTVTPNPEH 180
 DB 121 QTPMTSTDLTKQNTYVKVNHAPTYINTSPKSIKKATPKKSPKVAFTVTPNPEH 180
 QY 181 HYPDNRVEEEDSOQKEDSVPEPLIOHOKDPSQFNYSDEDTNASVPTPLHTTKPTFA 240
 DB 181 HYPDNRVEEEDSOQKEDSVPEPLIOHOKDPSQFNYSDEDTNASVPTPLHTTKPTFA 240
 QY 241 QLNKNNEVNSEPALTDMLKR 263
 DB 241 QLNKNNEVNSEPALTDMLKR 263

RESULT 3

AAE19799

ID AAE19799 standard; Protein: 1664 AA.

XX AAE19799;

XX 18-JUN-2002 (first entry)

XX *Candida albicans* integrin-like protein (alphaIntlp).

XX Integrin-like protein; alphaIntlp; gene; fungicide; contraceptive;
 XX HIV; human immunodeficiency virus; vaccine; cholera.

XX *Candida albicans*.

XX Key Location/Qualifiers

```

FT Misc-difference 251
FT /note= "Encoded by CTG"
FT Misc-difference 295
FT /note= "Encoded by CTG"
FT Misc-difference 339
FT /note= "Encoded by CTG"
FT Misc-difference 519
FT /note= "Encoded by CTG"
FT Misc-difference 1102
FT /note= "Encoded by CTG"
FT Region
FT 1149..1151
FT /note= "RGD site"
PN US6346411-B1.
PD 12-FEB-2002.
PF 08-MAR-1999; 99US-0264604.
PR 03-MAY-1996; 96US-0642846.
XX (MINU ) UNIV MINNESOTA.
XX Hostetter MK, Gale CA, Bendel CM, Tao N;
XX WPI; 2002-224995/28.
XX N-PSDB: AAD31519.
XX
XX Polynucleotide encoding Candida albicans protein with integrin-like
XX motifs and protein, useful as vaccines and for raising antibodies for
XX inhibiting adhesion of pathogen to cells preferably epithelial cells
XX
XX Claim 3; Column 15-18; 25pp; English.
XX
XX The present invention relates to an isolated polynucleotide encoding
XX Candida albicans protein with integrin-like motifs. Candida albicans
XX protein is utilised in vaccines (for gastrointestinal pathogens like
XX cholera) and as antigens to prepare anti-peptide antibodies, which are
XX utilised in inhibiting adhesion of C. albicans to cells preferably
XX epithelial cells. In the gastrointestinal tract, expression of spermicides
XX by S. cerevisiae transformed with the C. albicans integrin-like gene on
XX an extrachromosomal plasmid could provide a cheap and infrequent method
XX of contraception. Also, synthesis of protein-based antiretroviral
XX agents could help to reduce transmission of human immunodeficiency virus
XX (HIV) in the birth canal. The present sequence is Candida albicans
XX integrin-like protein (alphaIntlp).
XX
XX Sequence 1664 AA;
XX
XX Query Match 100.0%; Score 1386; DB 23; Length 1664;
XX Best Local Similarity 100.0%; Pred. No. 5e-107;
XX Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT 4
AAW99456
ID AAW99456 standard; protein; 236 AA.
XX
XX AAW99456;
AC 08-JUN-1999 (first entry)
XX
XX Amino acids 218-453 of C.albicans integrin-like protein 1.
XX
XX Integrin-like motif; vaccine; immune response; antibody; inhibition;
XX adhesion; endothelial cell; pathogenesis; infection.
XX
XX Candida albicans.
XX
XX US5886151-A.
XX
XX 23-MAR-1999.
XX
XX 03-MAY-1996; 96US-0642846.
XX
XX 03-MAY-1996; 96US-0642846.
XX
XX (MINU ) UNIV MINNESOTA.
XX
XX Bendel CM, Gale CA, Hostetter MK, Kendrick K, Tao N;
XX WPI; 1999-242618/20.
XX
XX New isolated Candida albicans protein with integrin-like motifs
XX
XX Claim 1; Column 35; 21pp; English.
XX
XX Peptides AAW99456-w99461 are derived from a Candida albicans protein
XX with integrin-like motifs, alpha-INP1. This sequence represents amino
XX acids 218-453 of alpha-INP1. The peptides can be used for producing
XX vaccines for stimulating an immune response. The antibodies can inhibit
XX the adhesion of C.albicans to cells, particularly endothelial cells.
XX This blocking activity of the adhesion to cells can reduce or prevent
XX subsequent events in the pathogenesis of invasive candidal infection.
XX
XX Sequence 236 AA;
XX
XX Query Match 17.4%; Score 241; DB 20; Length 236;
XX Best Local Similarity 100.0%; Pred. No. 2.8e-12;
XX Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


xx	Drosophila; developmental biology; cell signalling; insecticide;
xm	pharmaceutical.
xx	
os	Drosophila melanogaster.
pn	
xx	WO200171042-A2.
xx	
pd	27-SEP-2001.
pe	23-MAR-2001; 2001WO-US09231.
pr	23-MAR-2000; 2000US-191637P.
xx	11-JUL-2000; 2000US-0614150.
xx	
pa	(PEKE) PE CORP NY.
xx	
pt	Venter JC, Adams M, Li PWD, Myers EW;
xx	
dr	WPI; 2001-656860/75.
xx	N-PSDB; ABL02343.
xx	
pt	New isolated nucleic acid detection reagent for detecting 1000 or more
xx	genes from Drosophila and for elucidating cell signalling and cell-cell
xx	interactions -
ps	
xx	Disclosure; SEQ ID NO 1512; 21pp + Sequence Listing; English.
cc	
cc	The invention relates to an isolated nucleic acid detection reagent
cc	capable of detecting 1000 or more genes from Drosophila. The invention is
cc	useful in developmental biology and in elucidating cell signalling and
cc	cell-cell interactions in higher eukaryotes for the development of
cc	insecticides, therapeutics and pharmaceutical drugs. The invention
cc	discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
cc	sequences (AB101840-ABL16175) and the encoded proteins
cc	(ABB57737-ABB72072).
cc	
cc	The sequence data for this patent did not form part of the printed
cc	spectification, but was obtained in electronic format directly from WIPO
cc	at ftp.wipo.int/pub/published_pct_sequences.
cc	
xx	
xx	Sequence 767 AA:
xx	
yy	Query Match 9.6%; Score 133; DB 22; Length 767;
db	Best Local Similarity 24.1%; Pred. No. 0.014;
Matches	87; Conservative 31; Mismatches 121; Indels 122; Gaps 16,
oy	2 NSTPKLRIDKHSNLOLOPOSSASIFNS-----PTKPLNFPRKNRP-----SIDPNS 51
db	232 NSQPSVYNAOLANLAIIGDGASAATPVGSSNTSSNPMPPIORSNSPOSISGLSPSS 291
yy	52 SSDTYSFODDQEGK-----EEKRTAFOTSFDRNFDLNSIDIOO----- 92
db	292 ISNMSSSASCSCGGSGSLNGNNNGSTTTTALITS-----NCNVLDQLSAGNVY 343
oy	93 -----TIOHQOOP-----QQOOOLSQ-----TDNLNIDE---FSQTPTSTL--- 128
db	344 TLTNLPFOOHQOORPORQOQQOOLHQNMTATGNSNNNNIMRSTGLVSIKTNNTPAKSV 403
oy	129 -DLTONPTVDKV-----NEHN-----APTITNS--PKSITIK 159
db	404 SNLMAMATAASKILKLATNVANGSSMDIVDOOQHNTTTTGQVLVNSKRLPNLVNIG 463
oy	160 KATPKASPRK-----VAFTVNPEINHXY-----PDNVVEEEDSOOKR 197
db	464 ILTPTATSPKNTPTGTLETTTNGSTGOQAALVASLSKKTWILLDFETOLVOOVQOQOQO 523
oy	198 DSVERPL-----IQHWKDPDSQFNYSDEDTNASVPTRPYLHTTKPFAOLLKNNENNSE 252
db	524 QQQQQOOLQOQVVQGHNVILPTSTSSVSAGNSSKSSSPTP-----QQQQQDLKKKGKELLE 578
oy	253 P 253
db	579 P 579

ID	AAW48896	standard; Protein; 971 AA.
XX	AAW48896;	
XX	13-OCT-1998	(first entry)
XX	Candida albicans	
XX	Cac1A4; protein kinase; Ste20p family; screening; virulence;	
XX	hypthal formation; pathogenic fung1; inhibitor; inflammation;	
XX	antimycotic.	
XX	Candida albicans.	
XX	WO9818927-A1.	
XX	07-MAY-1998.	
XX	29-OCT-1997;	97WO-CA00809.
XX	30-OCT-1996;	96US-0029458.
XX	(CANA) NAT RES COUNCIL CANADA.	
XX	Leberer E, Thomas DY;	
XX	WPI: 1998-272222/24.	
XX	N-PSDB: AAV32554.	
XX	In vitro screening test for agents that inhibit Candida genes	
XX	involved in virulence - and transition to hyphal form, potentially	
XX	useful as antimycotic agents	
XX	Disclosure: Fig 7; 79pp: English.	
XX	The sequence is that of the Cac1A4 protein which can be used	
XX	in the development of an in vitro screening test for compounds	
XX	that inhibit biological activity of the protein and a system for	
XX	measuring its activity. The protein is involved in virulence and	
XX	hypthal formation. Inhibitors are potentially useful for rendering	
XX	pathogenic fungi (any species in which hyphal induction by kinase	
XX	occurs) avirulent and/or to treat inflammation.	
XX	Sequence 971 AA;	
XX	Query Match 9.6%; Score 133; DB 19; Length 971;	
XX	Best Local Similarity 24.3%; Pred. No. 0.019; Mismatches 85; Indels 78; Gaps 12;	
XX	Matches 63; Conservative 33; Mismatches 85; Indels 78; Gaps 12;	
XX	4 TESKLLPIDKSHILOPOSSASIT-----FNSPTK---PLN-----FPR 40	
XX	358 TEYHLTOLNGSSH--QHTSSSGSLPSSGNNNNNNSTNNNTKKNVSPLNANKSELLPA 414	
XX	41 TMSKPSLDNNSDPYTSQDOEKGEKKKOTAFOTSPFRNDDLNSIDIOQTIOHQOO 100	
XX	415 RRAPEPTSGTSDYTSNNHODRSGEYO-----RQORTPSOQOQOQOKHOYOO 465	
XX	101 PQOQOOLSGTDNNLIDEFSGFQPMT-----STLDLTQKNFTYKVNENHAPYINTSPNKS 156	
XX	466 KQOQOQOQOQ-----QOPLSHOGGSHIRKQV-----PTLPSSGP-- 502	
XX	157 IKKATPKASPKKVAFTYTNPEIHAYPDNRVEEDOSQOKEDSVPEPLIQHOKWD-PSQF 215	
XX	503 -TQAASGSGKSPSKI-----HPLDKIQGTNNYIKSSGTDANQYDAKQFIKPF 550	
XX	216 NY---SDEDTNASVPTTP 231	
XX	551 NLOLKKSQOOLASKQSPSP 569	

```

RESULT 9
AAG85039
ID   AAG85039 standard; Protein; 1174 AA.
XX
AC
XX
XX
DT   11-SEP-2001 (first entry)
XX
DE   Shrimp white spot Bacilliform virus (WSBV) protein 130.
XX
XX   Shrimp white spot Bacilliform virus; MSBV; diagnosis; viral
KW   antiviral agent; gene expression; antisense construct;
KW   transgenic viral resistant shrimp.
XX
XX   White spot syndrome virus.
OS
XX
XX   W0200138351-A2.
CN

```

PE 08-NOV-2000; 2000WO-0528888.
 XX
 PR 24-NOV-1999; 99CN-0124717.
 XX
 PA (PENY-) PE CORP NY.
 PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
 PA (TINO-) SINOGENOMAX CO LTD.
 XX
 PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
 XX
 DR WPI: 2001-355877/37.
 DR N-PSDB; AAH62819.
 XX
 XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus
 PT (WSBV), useful for producing viral polypeptides that can be used to
 PT screen for agents that are useful for treating WSBV infection -
 XX
 PS Claim 1; Figure 3; 626pp; English.

CC The invention provides the primary nucleotide sequence of the MSB1 genome
CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC encoded proteins (MAG64910-MAG65051) and oligonucleotide sequences
CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC molecules and proteins of the invention are useful for diagnosis and
CC monitoring viral infection, in screens for antiviral agents and for
CC monitoring viral gene expression or activity during a treatment regimen.
CC The nucleic acid molecules are also useful as antisense constructs to
CC control viral gene expression in infected cells and tissues and to create
CC transgenic viral resistant shrimp

SO	sequence	1174	AA;
	Query Match	9.5%;	Score 131.5; DB 22; Length 1174;
	Best Local Similarity	20.4%;	Pred. No. 0.033;
Matches	58; Conservative	51;	Mismatches 106; Indels 69; Gaps 11.

[illegible]

Qy 204 LIQHWKPPSQNYSDEDTASVPTPLHTTKPTFAQLINKN 247
 || | : || | : |
 Db 466 PPHQ---PQL---KPTNILLPPPPPTNQSLFSN---NNNN 500

RESULT 10	
ABB68896	
ID	ABB68896 standard; Protein; 2439 AA
EV	

DT 26-MAR-2002 (first entry)
XY

DE *Drosophila melanogaster* polypeptide SEQ ID NO 33480.
YV

KW Drosophila; developmental biology; cell signalling; insecticide; abamectin; cell

KW pharmaceutical.

OS *Drosophila melanogaster*.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

[illegible]

(F B I)
XX XX

FT VEHICLE CO, AUDUBON M,
XX

DR WPL; 2001-030880/
DR N-PSDB; ABL12999.

XX
PT
New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from L
PT interactions

PS- XX Disclosure: SEO ID NO 33480: 21mb + Sequence Listing: English

XX The invention relates to an isolated nucleic acid detection reagent

capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signaling

cell-cell interactions in higher eukaryotes for the development of

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC (ABB57737-ABB72072).

CC specification, but was obtained in electronic format directly from WIPO

XX

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Query Match:	9.5%:	Score	131.5:	DB	22:	Length	2439:
Best Local Similarity	19.9%:	Pred	NO.0.085:				
Matches	63:	Conservative	38:	Mismatches	109:	Indels	107:
						Gaps	11:

```
0Y      10 PIDKSHLQL-----QPOSSASIFNSPTKPLNFPRP-----NSKPSL----- 47
```

00 -----DPNSSPTYTSSEDOEKSGREEKKDTAFOTSEDRNEDJDN----- 866
 01
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 00
 01
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 00
 01
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 00
 01
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 00
 01
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 00
 01
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 8

Db 1065 ÖGVKERAKDPTAPKÖKREBERAKGNVKTKAALKEEIKLRLKLEKVEYTPPAAGAGYEGÖIII 1124

```

87 -----SIDIQQTIGHQRRQRPQQQQQLSQTDDNNLIDFSFQT 122
QY

```

Db 1.125 N0Q0Q0QTI VVN NHM Q0L P M Q T Q L L A L P T Q Q Q Q Q Q Q Q Q Q Q Q Q Q M Q R P Q M I S Q Q H Q Q Q 1.184

QY 123 PMSTL-----DLTKONPTVD-----KVENHAPTYIN-----TSPNKSI 157

This Page Blank (uspto)

Shears, Beverly

From: Devi, Sarvamangala
Sent: Friday, March 14, 2003 4:38 PM
To: Shears, Beverly
Subject: 09/964,858

Hello Beverly:

Would you please perform a sequence and an interference search for SEQ ID NO: 1 and a fragment comprising amino acid residues 1-263 of SEQ ID NO: 1 in case 09/964,858?

Thanks.

S. DEVI, Ph.D.
AU 1645
CM1-7E15

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 1E05 Tel: 308-4994

This Page Blank (uspto)

GenCore version 5.1.4-P5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:21:44 ; Search time 4.09445 Seconds

(without alignments)
1889.932 Million cell updates/sec

Title: US-09-964-858-1_COPY_1_263

Sequence: 1 MNSTPSKLPIDKSHLQLO.....NKNNEVSEPEALTDMLKR 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Optical number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1386	100.0	1664	US-09-599-652-2	Sequence 2, Appli
2	1386	100.0	1664	US-08-642-846-2	Sequence 2, Appli
3	1386	100.0	1664	US-09-264-604-2	Sequence 2, Appli
4	241	17.4	236	US-09-599-652-3	Sequence 3, Appli
5	241	17.4	236	US-08-642-846-3	Sequence 3, Appli
6	241	17.4	236	US-09-264-604-3	Sequence 3, Appli
7	127.5	9.2	1335	US-09-134-001C-3716	Sequence 3716, Ap
8	118	8.5	10182	US-09-134-001C-3159	Sequence 3159, Ap
9	114	8.2	1164	US-08-923-992A-10	Sequence 10, Appl
10	112.5	8.1	565	US-08-961-083-218	Sequence 218, App
11	112	8.1	984	US-08-242-932-2	Sequence 2, Appli
12	112	8.1	984	US-08-714-481-2	Sequence 2, Appli
13	112	8.1	984	PCT-US95-06111-2	Sequence 2, Appli
14	111	8.0	652	US-09-134-001C-3517	Sequence 3517, Ap
15	111	8.0	699	US-09-134-001C-4054	Sequence 4054, Ap
16	110.5	8.0	903	US-08-853-310-2	Sequence 2, Appli
17	110	7.9	413	US-08-700-749A-1	Sequence 1, Appli
18	110	7.9	413	US-09-020-684-1	Sequence 1, Appli
19	110	7.9	413	US-09-020-467-1	Sequence 1, Appli
20	110	7.9	413	US-09-020-685-1	Sequence 1, Appli
21	110	7.9	413	US-09-020-683-1	Sequence 1, Appli
22	109.5	7.9	1848	US-08-296-791-6	Sequence 6, Appli
23	109.5	7.9	1848	PCT-US95-10661A-6	Sequence 6, Appli
24	109	7.9	1104	US-08-923-992A-4	Sequence 4, Appli
25	109	7.9	1128	US-08-923-992A-6	Sequence 6, Appli
26	107.5	7.8	715	US-09-215-221-25	Sequence 25, Appli
27	107.5	7.8	1098	US-08-923-992A-8	Sequence 8, Appli

28	107	7.7	396	US-08-430-024-2	Sequence 2, Appli
29	107	7.7	396	US-08-782-009-2	Sequence 2, Appli
30	107	7.7	396	US-09-017-302-2	Sequence 2, Appli
31	107	7.7	1164	US-08-923-992A-2	Sequence 2, Appli
32	107	7.7	1702	US-08-296-791-5	Sequence 5, Appli
33	107	7.7	1702	PCT-US95-10661A-5	Sequence 5, Appli
34	105.5	7.6	1541	US-08-296-791-3	Sequence 3, Appli
35	105.5	7.6	1541	PCT-US95-10661A-3	Sequence 3, Appli
36	105.5	7.6	1545	US-08-296-791-4	Sequence 4, Appli
37	105.5	7.6	1545	PCT-US95-10661A-4	Sequence 4, Appli
38	104.5	7.5	219	US-09-134-001C-5651	Sequence 5651, Ap
39	104	7.5	788	US-08-918-914-4	Sequence 4, Appli
40	104	7.5	1099	US-09-442-100-2	Sequence 2, Appli
41	102.5	7.4	544	US-08-935-760-2	Sequence 2, Appli
42	102.5	7.4	664	US-08-421-661-6	Sequence 6, Appli
43	102	7.4	351	US-08-929-417-2	Sequence 2, Appli
44	102	7.4	930	US-09-134-001C-5314	Sequence 5314, Ap
45	101.5	7.3	542	US-07-814-964-13	Sequence 13, Appli

ALIGNMENTS

RESULT 1
US-09-599-652-2
Sequence 2, Application US/09599652
Patent No. RE37741
GENERAL INFORMATION:
APPLICANT: HOSTETTER, MARGARET K.
APPLICANT: GALE, CHERL A.
APPLICANT: BENDEL, CATHERINE M.
APPLICANT: TAO, NIAN-JUN
APPLICANT: KENDRICK, KATHLEEN
TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: MUEITING, RAASCH, GERHARDT & SCHWAPPACH, P. A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/599,652
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/642,846
FILING DATE: 03-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: MUEITING, ANN M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110.00280101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1664 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-599-652-2
Query Match 100.0%; Score 1386; DB 1; Length 1664;
Best Local Similarity 100.0%; Pred. No. 2.8e-115;

	Matches	263	Conservative	0	Mismatches	0	Indels	0	Gaps	0
QY	1	MNSTPSKLLPIDKSHHLQLOPOSSSASIFNSTPKPLNPRTMSKSPLOPNSSDRTYSIQ								60
Db	1	MNSTPSKLLPIDKSHHLQLOPOSSSASIFNSTPKPLNPRTMSKSPLOPNSSDRTYSIQ								60
QY	61	DOEKREEEKKDTAFOTSFDRNFDLNSIDIOQTIOHQOQOPQOQOOLSTQDNLLIDEFSEF								120
Db	61	DOEKREEEKKDTAFOTSFDRNFDLNSIDIOQTIOHQOQOPQOQOOLSTQDNLLIDEFSEF								120
QY	121	QTPMSTLDTLRKQNTFYVKNVNNHAPTYINTSPNKSIMKKATPRKSPKKVATVYNPEIH								180
Db	121	QTPMSTLDTLRKQNTFYVKNVNNHAPTYINTSPNKSIMKKATPRKSPKKVATVYNPEIH								180
QY	181	HYPDNRVEEEDSOQEKDSVEPPLIQHOQKPSQSNYSDEDTNAAVPTPLHTTKPTFA								240
Db	181	HYPDNRVEEEDSOQEKDSVEPPLIQHOQKPSQSNYSDEDTNAAVPTPLHTTKPTFA								240
QY	241	QLLNKNNEVNSEPALTDMLKLR								263
Db	241	QLLNKNNEVNSEPALTDMLKLR								263

```

RESULT 2
US-08-642-846-2
: Sequence 2, Application US/08642846
: Patent No. 5866151
:
: GENERAL INFORMATION:
: APPLICANT: HOSTETTER, MARGARET K.
: APPLICANT: GALE, CHERYL A.
: APPLICANT: BENDEL, CATHERINE M.
: APPLICANT: TAO, NIAN-JUN
: APPLICANT: KENDRICK, KATHLEEN
:
: TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
: TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
:
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MUETING, RAASCH, GERHARDT & SCHWAPPACH, P.A.
: STREET: 119 NORTH FOURTH STREET, SUITE 203
: CITY: MINNEAPOLIS
: STATE: MINNESOTA
: COUNTRY: USA
: ZIP: 55401
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/642,846
: FILING DATE: 03-MAY-1996
: CLASSIFICATION: 424
:
: ATTORNEY/AGENT INFORMATION:
: NAME: MUETING, ANN M.
: REGISTRATION NUMBER: 33, 977
: REFERENCE/DOCKET NUMBER: 110, 00280101
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-305-1217
: TELEFAX: 612-305-1228
:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1664 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-08-642-846-2
:
: Query Match 100.0%; Score 1386; DB 2; Length 1664;
: Best Local Similarity 100.0%; Pred. No. 2.8e-115;
: Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0
:
: 1 MNSPTSKLLPIDKSHLQLOPSSASATFNSPTKPLNFPRTNSKPSLDPNSSDITYTSEQ 600

```

Db	1	MMSFSLKLLPIDKSHSLDLPQSSASSIFNSPTKPLNEPRTNSKPSLDPNSSSDPTYSSEQ	60
QY	61	DOEKKEEKKQTATQTSRDRNFIDLNSIDIQOTIQHOQOQPPQOQOOLQSDTNNILDEFSF	120
Db	61	DOEKKEEKKQTATQTSRDRNFIDLNSIDIQOTIQHOQOQPPQOQOOLQSDTNNILDEFSF	120
QY	121	QTPMTSTLDLTQNPYDVKVVENHAFTYINTSPKSIIMKKATPKASPKKAAFTVTNDEIH	180
Db	121	QTPMTSTLDLTQNPYDVKVVENHAFTYINTSPKSIIMKKATPKASPKKAAFTVTNDEIH	180
QY	181	HYPDNRVEEEDQSOQKEDSVPEPLIQHQMKPDSQFNYSDEDTNNSVPTPPLHTTKPTFA	240
Db	181	HYPDNRVEEEDQSOQKEDSVPEPLIQHQMKPDSQFNYSDEDTNNSVPTPPLHTTKPTFA	240
QY	241	QLLNKNNEVNSEPEALDMMKLKR	263
Db	241	QLLNKNNEVNSEPEALDMMKLKR	263

```

RESULT 3
US-09-264-604-2
: Sequence 2, Application US/09264604
: Patent No. 6346411
: GENERAL INFORMATION:
: APPLICANT: HOSTETTER, MARGARET K.
: APPLICANT: GALE, CHERYL A.
: APPLICANT: BENDEL, CATHERINE M.
: APPLICANT: TAO, NIAN-JUN
: APPLICANT: KENDRICK, KATHLEEN
: TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
: TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MUETING, RAASCH, GERHARDT & SCHNAPPACH, P.A.
: STREET: 119 NORTH FOURTH STREET, SUITE 203
: CITY: MINNEAPOLIS
: STATE: MINNESOTA
: COUNTRY: USA
: ZIP: 55401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/264,604
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/642,846
: FILING DATE: 03-MAY-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: MUETING, ANN M.
: REGISTRATION NUMBER: 33,977
: REFERENCE/DOCKET NUMBER: 110.00280101
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-305-1217
: TELEFAX: 612-305-1228
: INFORMATION FOR SEQ ID NO. 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1664 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-264-604-2
:
: Query Match 100.0%; Score 1386; DB 4; Length 1664;
: Best Local Similarity 100.0%; Pred. No. 2,8e-115;
: Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
1 MNSTPSKLLPDIKHSLLQLOPSSASIFNSPTKPLNFPRTNSKPSLDNSSSDITYTSEQ 60

```

Db 1 MNSPRLKLPIDKSHLQLOPSSASIFNSPTKPLNEPRINSKSLDPNSSDPTYSQ 60
Qy 61 DQEGKEKKDTAFQTSFRNFDLDNSIDIOQTIOHQOQOQOQOOLSDTNNLIDESF 120
Db 61 DQEGKEKKDTAFQTSFRNFDLDNSIDIOQTIOHQOQOQOQOOLSDTNNLIDESF 120
Qy 121 QTPMTSLDLTKONPTVDKVENHAPTYINTSPKSIKKATPKASPKVAFTVTNPEIH 180
Db 121 QTPMTSLDLTKONPTVDKVENHAPTYINTSPKSIKKATPKASPKVAFTVTNPEIH 180
Qy 181 HYPDNREEEDQSOQKEDSVPEPLIOHQMDPSQFNTSDEDTNMSVPEPLHTTKPTFA 240
Db 181 HYPDNREEEDQSOQKEDSVPEPLIOHQMDPSQFNTSDEDTNMSVPEPLHTTKPTFA 240
Qy 241 QLLKNNVNSEPALDMLKR 263
Db 241 QLLKNNVNSEPALDMLKR 263

RESULT 4

US-09-599-652-3
; Sequence 3, Application US/09599652
; Patent No. RE37741
; GENERAL INFORMATION:
; APPLICANT: HOSTETTER, MARGARET K.
; APPLICANT: GALE, CHERYL A.
; APPLICANT: BENDEL, CATHERINE M.
; APPLICANT: TAO, NIAN-JUN
; APPLICANT: KENDRICK, KATHLEEN
; TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
; TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MUEITING, RAASCH, GERHARDT & SCHWAPPACH, P.A.
; STREET: 119 NORTH FOURTH STREET, SUITE 203
; CITY: MINNEAPOLIS
; STATE: MINNESOTA
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/599,652
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/642,846
; FILING DATE: 03-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MUEITING, ANN M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00280101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; OTHER INFORMATION: amino acid positions 218-453 from SEQ ID NO:2
US-09-599-652-3

Query Match 17.4%; Score 241; DB 1; Length 236;
Best Local Similarity 100.0%; Pred. No. 4.1e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 SDEDTNASVPPPLHTTKPTFAQLLNKNNVNSEPALDMLKR 263
Db 1 SDEDTNASVPPPLHTTKPTFAQLLNKNNVNSEPALDMLKR 46

RESULT 5

US-08-642-846-3
; Sequence 3, Application US/08642846
; Patent No. 5886151
; GENERAL INFORMATION:
; APPLICANT: HOSTETTER, MARGARET K.
; APPLICANT: GALE, CHERYL A.
; APPLICANT: BENDEL, CATHERINE M.
; APPLICANT: TAO, NIAN-JUN
; APPLICANT: KENDRICK, KATHLEEN
; TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
; TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MUEITING, RAASCH, GERHARDT & SCHWAPPACH, P.A.
; STREET: 119 NORTH FOURTH STREET, SUITE 203
; CITY: MINNEAPOLIS
; STATE: MINNESOTA
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,846
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: MUEITING, ANN M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00280101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; OTHER INFORMATION: amino acid positions 218-453 from SEQ ID NO:2
US-08-642-846-3

Query Match 17.4%; Score 241; DB 2; Length 236;
Best Local Similarity 100.0%; Pred. No. 4.1e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 218 SDEDTNASVPPPLHTTKPTFAQLLNKNNVNSEPALDMLKR 263
Db 1 SDEDTNASVPPPLHTTKPTFAQLLNKNNVNSEPALDMLKR 46

RESULT 6

US-09-264-604-3
; Sequence 3, Application US/09264604
; Patent No. 6346411
; GENERAL INFORMATION:
; APPLICANT: HOSTETTER, MARGARET K.
; APPLICANT: GALE, CHERYL A.
; APPLICANT: BENDEL, CATHERINE M.
; APPLICANT: TAO, NIAN-JUN
; APPLICANT: KENDRICK, KATHLEEN
; TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE

TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUETING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/264,604
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/642,846
FILING DATE: 03-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: MUETING, ANN M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110,00280101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 236 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
OTHER INFORMATION: amino acid positions 218-453 from SEQ ID NO:2
US-09-264-604-3

Query Match 17.4%; Score 241; DB 4; Length 236;
Best Local Similarity 100.0%; Pred. No. 4,1e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SDEDTNNAVPPPTPLHTTKPTFAQLLNKNNEVSEPALTDMLKLR 263
|||||
DB 1 SDEDTNNAVPPPTPLHTTKPTFAQLLNKNNEVSEPALTDMLKLR 46

RESULT 7
US-09-134-001C-3716
Sequence 3716, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3716
LENGTH: 1335
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3716

Query Match 9.2%; Score 127.5; DB 4; Length 1335;
Best Local Similarity 21.2%; Pred. No. 0.006;
Matches 54; Conservative 41; Mismatches 125; Indels 35; Gaps 7;

QY 18 QLOPOSSASIFNSPTKPLNFPRTNSKPSLDPNSSDVTY-SEDDQEKGEKKDTAFQT 76
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 32 QPOQNSHKRWLVDDQTLKQAEKAKSEVLTOSTNVSTQYQDPYQAPQDPQSTTYDA 91
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 77 SPDRNFDLNSIDIOQTIOHQOOOPQOOQOLSDTNLIDFES-OTPMSTLDTKQNP 135
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 92 SLD-----EMSTYNEISSNQOQSLSTDANONOTNSVTKNOQEEITDLTQEDK 140
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 136 TVDKVNEHAPTYINTSPNSIKKATPKKASPKKVAFTVTPNPEIHHPDMRVEEED-QSQ 194
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 141 TSDITNLOETQSYAKENEDLGNANNEQDKM--TASQPSENALETQTASNDNESQ 198
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 195 QKEDSV-----EPPLIOHQWKPDSQFNSDEDTNNAVPPPTPLHTTKPTFAQLLNKN 247
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 199 QKSQVYSEQNENATPKVSNMTNAGVNFYDDEDDSDSTHLEPISL-----N 246
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 248 EVNSEPEALTDMLKLR 262
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 247 NVNATSKOTTSYKXK 261
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
RESULT 8
US-09-134-001C-3159
Sequence 3159, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3159
LENGTH: 10182
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 8.5%; Score 118; DB 4; Length 10182;
Best Local Similarity 21.2%; Pred. No. 0.67;
Matches 61; Conservative 45; Mismatches 112; Indels 70; Gaps 10;

QY 11 IDKHSHLQLOPOSSASIFNSPTKPLNFPRTNS----KPSLDPNSSDVTYSEDDQEKGK 66
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 4526 INNESHILNNAQKDNFKAQVNSA-----PNHNTLETIKNRADTLNOSMTALSISIADEVN 4579
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 67 EEKKDTAFQTSFDRNFDLNSIDIOQTIOHQOOOPQOOQOLSDTNLIDFESFQPMTS 126
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 4580 QKOQENTLDSNNKRQDYDVAANAKILNQTOSPTMSA-----DIVDQKADYKRTK 4632
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 127 T-----IDLTKQNP-----TVDKVNEHAPTY--INTSPNSIKKATPKKASPKVA 171
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 4633 TALDGNORLEVAQOALNHLNTLNDLDAQRQTLTITINSPIINSVNAKEKANTVNTA 4692
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 172 FYTYNPEIHHYP-----NRVEEDSOQKEDSVPEPLIOHQWKP 212
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 4693 MTQLKQTIANYDELHDGNYINADKDKKAYNNAVNNAKQLINQSDANOQL-----DP 4746
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 213 SDFNYSDEDTNNAVPPPTPLHTTKPTF--AQLLNKNNEVSEPALT 257
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 4747 AEIN-----KYTORVNTTKNDLNGNDKLALEAKRDANTTIDGLT 4784
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
RESULT 9
US-08-923-992A-10
Sequence 10, Application US/08923992A
Patent No. 6280738

```

; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-1ga Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438, 0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-923-992A-10

Query Match      8.2%; Score 114; DB 4; Length 1164;
Best Local Similarity 18.5%; Pred. No. 0.08;
Matches 50; Conservative 52; Mismatches 80; Indels 88; Gaps 11;

OY 22 OSSASIFNSPTKPLNFPRTNSKPSLDP-----NSSDYTSBDOCKEKKD--- 71
DB 64 OGNNSSSELETKKIEIPTDIDKKAVERPLEKTAGETSAITDTGKRREKOLQOQNNLNKNVH 123
OY 72 -----TAFQTSFDRNFIDNSIDIQOTI-----OHQO---QQPOOQOOLS 108
DB 124 NTILSHQKNEFKTIDETNDSDALLEENQFETNRLHLIKQHEVEKDKKAKOQKTLK 183
OY 109 QTD-----NLIDERSFQTPMTSTL-----DLTKONPVTD----- 138
DB 184 QSDTVDLSNIDKELNHOKSQVETMAEOLGITNEDEKDSMLKKIETDRKQAOQADKKEDAE 243
OY 139 -KVNENHAPTYINTSP-----NKSIMKATPKASPKKAVFYVTNNEIHYPDN----- 185
DB 244 KVRBELGKLTSTKAGLDQETQIEQENHKETTSSENTQKVD-----BHYPNSLONLAQ 295
OY 186 -RVEEDSOQKEDSEVPEPLIQHOMKDPQO 214
DB 296 KSLSELDKATNEQATQ---VKNQPLENMQ 322

RESULT 10
US-08-961-083-218
; Sequence 218, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 218:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-218

Query Match      8.1%; Score 112.5; DB 4; Length 565;
Best Local Similarity 22.0%; Pred. No. 0.041;
Matches 61; Conservative 37; Mismatches 102; Indels 77; Gaps 13;

OY 1 MNSTPSKLLPIDKSHLQLOPOSSA--SIFNSPTKPLNFPRTNSKPSLDPNSSSTPTYS 58
DB 78 VSKPEKVAKVPES---QPSDKPAEESKVEQAGEPV-APREDEKAPVPE----- 123
OY 59 EDOEKEKREKKDTAFQTSFDRNFIDNSIDIQOTIHOQOQOPOO-----OOLSQTDNN 113
DB 124 EKQPEAREBEKA-----VEETRKQESPTDITAEETVKEKETVQ 164
OY 114 LIDFSFQTPMTSTLDLTKONPVTDKVNNENHAPTYINTSPNKSIMKATPKASPKKVAFT 173
DB 165 SIEQPVETPAVEKQTEPEEPKVEQAGEPVAPREDEQAP-----TAPVEPEK----- 212
OY 174 VTNPELIHYPDRNVEEDSOQKEDSV---EPLIQHOMKDPQOFNYSDEDTNASTPPT- 229
DB 213 --QPEVPE--EKAVAETTPPEDEKIGIKTPE-----VDKSELN-NOIDKASVSPTD 261
OY 230 -----PLHTTKPTPEAQLLNKNNEVNSPEAL 256
DB 262 YSTASYNALGPVLETKAGVYASEPVQPEVNSSTNKL 298

RESULT 11
US-08-242-932-2
; Sequence 2, Application US/08242932
; Patent No. 5595740
; GENERAL INFORMATION:
; APPLICANT: Brady, L. Jeannine
; TITLE OF INVENTION: Cloning of No. 5595740-1ga Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
```


STREET: 2421 N.W. 41st Street, Suite A-1
City: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06111
FILING DATE:
CLASSIFICATION:
Prior Application DATA:
APPLICATION NUMBER: US 08/242,932
FILING DATE: 16-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Salliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UFI142
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEO ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 984 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-06111-2

Query Match	8.18;	Score 112;	DB 5;	Length 984;
-------------	-------	------------	-------	-------------

Matches 56; Conservative 54; Mismatches 115; Indels 88; Gaps 11;

```

0Y 22 QSSASJINSPTKPLNFPRTSKSLDP-----NSSDPYTSQOQKGEKKRQ---71
Db 27 QGNSSSSSELTTKMEIPTDITDKKAVEVEVETAGETSATDQGRKKRKLQOKNNKLKMDVD 86
0Y 72 -----TAFQTSFDRNFDLNSIDIQDT-----QHQQ---QOPQOQOOLS 108
Db 87 NTLISHEOKNEFKRIKETNDSDLLELENOFNETNRLHLITKHOEEKDKKAKQOOKTLK 146
0Y 109 QTD-----NNLIDFSEFQTPMTSL-----DLTKONPTVDKVENH 144
Db 147 QSDRFVDSLNSIDKELNHQKQSEVKMAEQKGTINEDKSMULKIEDIKKQAOQADKKRDAE 206
0Y 145 APTIYNTSPNKSINKKATPKKASPKKVAFTYVNPETLHHYPDNRVDEE-----DOSQOK-EDS 199
Db 207 VKVOLEEBEASHKLQAVEDFRKKRKTSEQV-----PKRVRKRDILAANNENQOKTELT 259
0Y 200 VEPPLIOHQWMDPSQOFNYSDEDTNAAVPPPLPLHTT-----KPFQAQLLNKNEV 249
Db 260 VSPENIT-----VEGEDVAFVYAKAKSDKTTLIDFSDLTKNYPNSVSDRISTRYKT 310
0Y 250 NSEPEALTDMLK 262
Db 311 NTDNKKIAETITK 323

```

RESULT 14
US-09-134

```

: Sequence 3511, Application US/09134001C
: Patent No. 6380370
:
: GENERAL INFORMATION:
:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964

```

```

: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3517
: LENGTH: 652
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-3517

```

Query Match	8.08; Score 111; DB 4; Length 652;
Query Match	8.08; Score 111; DB 4; Length 652;

Matches 49; Conservative 26; Mismatches 57; Indels 52; Gaps 10;

```

QY 12 DKSHILOPOSSASINSPTKPLNPRNTSKS-----LDNNSSDPYTSODD-EK 64
Db 94 ENOSGNVOKKSNQIODDSTSTSPINDOKOTSMEOQSKDNHVTNRSODTY PKGONODDK 153
QY 65 GKEEKRD-----TAFQ-TSPDRNFDLNSIDIOOTIGHOO-----OPQOOOOLSOYD 111
Db 154 GKQOFKEDNQHSGTGNQRTQNGNNDQOSS-DKQHNPSQDQDSSSKGQFQOSIEDRO 212
QY 112 NNLDIEFSQTRMTSTEDLTQONPR--VDKYNENHAPRTIYNTSPKSI--MKATPKRAS 166
Db 213 K-----TVKOPSSKVNKIG-----NTKTDKQIVTKNOKKOTSLTS 246
QY 167 PKKV 170
Db 247 PRVV 250

```

RESULT 15

```

US-09-134-001C-4054
; Sequence 4054, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4054
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4054

```

Query Match 8.08; Score 111; DB 4; Length 699;

Matches 53; Conservative 37; Mismatches 103; Indels 38; Gaps 10;

```

OY 11 IDKSHILOLOPOSSAS1-----FNSPPKPLNFPRTNKPSLDRNRSSDYYTSRDODEKG 65
Db 134 YAKHS--BEKPOOEVELEKHAASENNOTLHSAKAOSNBDVKTGPSOLNTPTAAGEDSOKE 191
OY 66 KEERKDAFQTSFD--RNFEDLNSIDIOQTIO-HOQOQPOQOQOLOSOTDNMLIDEFSPT 122
Db 192 NLSKRDQTSKTTDLLRKTGQNSKDSOSTEEVKEVKNPDYQYAKKDDDRVETFINLS 255
OY 123 PMTSTLDLTQK-NPTVDKVNENHAEPTYINTSPNKSIMKATPKASPKKVAETVTPETIH 181
Db 252 -KEEPLTKVDKQANPPTTK-----DKSKMKGSHDGLANLESNAVAT 293
OY 182 YDNNKVEEEDSOQKEDSEVPLLOHQK--DP-----SQFWSEDEPTNASY 226
Db 294 --NNSQOQVESEKNEEDTNKSARKOYKRNNDPILVLHGFGTDDINPSY 341

```

Mon Mar 17 12:50:24 2003

us-09-964-858-1_copy_1_263.ra1

Page 8

Search completed: March 17, 2003, 12:27:47
Job time : 17.0944 secs

FT	CARBOHYD	638	638	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	670	670	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	691	691	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	817	817	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	838	838	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	841	841	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	1018	1018	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	1082	1082	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	1100	1100	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	1103	1103	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	1200	1200	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	1571	1571	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	1593	1593	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	DOMAIN	95	106	POLY-GLN.	(POTENTIAL) .
FT	DOMAIN	283	286	POLY-ASN.	.
FT	DOMAIN	1283	1292	POLY-HIS.	.
FT	DOMAIN	1651	1660	POLY-GLN.	.
SO	SEQUENCE	1664 AA;	187859 MW; BAZEFD0DCB196790 CRC64;		
<hr/>					
Query Match:		100.0%;	Score 1386;	DB 1;	Length 1664;
Best Local Similarity:		100.0%;	Pred. No. 4,2e-77;		
Matches	263;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
OY	1	MNSTPSKLPIDPKHSHLQLOPOSSASIFNSPTKPLNPRTNSKSDSLDNSSDDYTTSQ	60		
Dd	1	MNSTPSKLPIDPKHSHLQLOPOSSASIFNSPTKPLNPRTNSKSDSLDNSSDDYTTSQ	60		
OY	61	DOEKKEKKDPAPFOTSPFRNFDLNDSIDIQTIOHQOQQPOQOOQLSOTDNNLIDEFSF	120		
Dd	61	DOEKKEKKDPAPFOTSPFRNFDLNDSIDIQTIOHQOQQPOQOOQLSOTDNNLIDEFSF	120		
OY	121	QTPMTSTLDLTQONPTVDKVNENHAPTYINTSPNKSIMKATPKASPCKVAFTVNPETH	180		
Dd	121	QTPMTSTLDLTQONPTVDKVNENHAPTYINTSPNKSIMKATPKASPCKVAFTVNPETH	180		
OY	181	HYBDNRVEEEDSOOKESVPEPLLQHOMKDDSOFRNYSBEDTNASVPRLPHHTTKPTPA	240		
Dd	181	HYPDRNVEEEDSOOKESVPEPLLQHOMKDDSOFRNYSBEDTNASVPRLPHHTTKPTPA	240		
OY	241	QLLNKNNEVNSEPALTDMKLKR	263		
Dd	241	QLLNKNNEVNSEPALTDMKLKR	263		
<hr/>					
RESULT 2					
CLIA4_CANAL	STANDARD:	PRT:	971 AA.		
<hr/>					
NO	CLIA4_CANAL	014427;			
DT	15-JUL-1998	(Rel. 36, Created)			
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DE	15-JUN-2002	(Rel. 41, Last annotation update)			
SN	Serine/threonine-protein kinase CLIA4 (EC 2.7.1.-).				
CC	Candida albicans (Yeast).				
OC	Eukaryota: Fungi: Ascomycota: Saccharomycotina: Saccharomycetes;				
OX	Saccharomycetales; mitosporic Saccharomycetales; Candida.				
RN	NCBI_TaxID=5476;				
RP	[1]				
RX	SEQUENCE FROM N.A.				
RA	MEDLINE=97411146; PubMed=9259554;				
RA	Leberer E., Zieglerbauer K., Schmidt A., Marcus D., Dignard D., Ash J.,				
RT	Johnson L., Thomas D.Y.:				
RT	"Virulence and hyphal formation of Candida albicans require the				
RL	Ste20p-like protein kinase Caclaf4."				
CC	Curr. Biol. 7:533-546(1997).				
CC	- FUNCTION: ESSENTIAL FOR VIRULENCE AND MORPHOLOGICAL SWITCHING				
CC	(HYPHAL FORMATION) OF C.ALBICANS.				
CC	- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.				
CC	STE20 SUBFAMILY.				
CC	- SIMILARITY: CONTAINS 1 CRIB DOMAIN.				
CC	- SIMILARITY: CONTAINS 1 PH DOMAIN.				
<hr/>					

CC This SWISS-PROT entry is copyright © It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

[illegible]

RESULT 3
LIPB_MYCPU STANDARD: PRT: 236 AA.
ID LIPB_MYCPU
AC 0918P6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipoprotein B precursor.
GN LIPB OR MYPV 5220.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
ON NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT;
RX MEDLINE=20245550; PubMed=10781561;
RA Shen X., Gunnar J., Yu H., French C.T., Zou N., Dybvig K.;
RT "Gene rearrangements in the *usa* locus of *Mycoplasma pulmonis*."
RL J. Bacteriol. 182:2900-2908(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
Mycoplasma pulmonis."
RL Nucleic Acids Res. 29:2145-2153(2001).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
(Potential).
CC -1- SIMILARITY: BELONGS TO THE PULMONIS LIPAB LIPOPROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC
CC EMBL; AF198037; AAF70133.1;
DR EMBL; AL445564; CAC13695.1;
DR Mypulist; MYPV_5220;
KM Lipoprotein; Membrane; Signal; Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 236 LIPOPROTEIN B.
FT LIPID 28 28 N-ACYL DIGLYCERIDE (POTENTIAL).
FT DOMAIN 82 158 ASD-RICH.
FT SEQUENCE 236 AA; 26661 MW; 183D1B968DCD1D3A CRC64;
Query Match 10.2%; Score 141.5; DB 1; Length 236;
Best Local Similarity 21.3%; Pred. No. 0.021;
Matches 47; Conservative 49; Mismatches 76; Indels 49; Gaps 7;
OY 52 SSDVTYSEDOEKGKEKEDKDTAFQTSFDRNFEDLNSIDIDIOOTIOHQOQOPO 103
DB 26 ASCVYNLAKEKDKOKESTNLSPEPKSMTSKTNFODKKDSTNKIDSESSSTQSQNTSE 85
OY 104 QOOLQSDTDNNLIDFSEFQTPMTSTLDTLKONPTVDKYN--ENHAPYINTSPNKSIMK 160
DB 86 SNQMTKVDSS-----KTNNLATQNNPKSSNVIQETNETKQONINPNNAVISE 135
OY 161 ATPKASPKKVAFTYNPEIHYPDNKRVEEEOOQKEDSVPEPLLOHOMKDPQSNYDSE 220
DB 136 --KQAKNDKNSLSNKOI--NNTLNKNDKTKQEND-----QKQESK 174
OY 221 DTNASVPTPLH-----TTKPTFAOLLKNNNEVNSEPE 254
DB 175 DSNFTSPKPTTHDPINKVISSQSTRLEMKNDQSNSESE 215
RESULT 4

META_PENCH
ID META_PENCH STANDARD: PRT: 499 AA.
AC 001870;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein META.
GN META.
OS Penicillium chrysogenum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
ON NCBI_TaxID=5076;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Miscosin 54-1255;
RX MEDLINE=94359480; PubMed=8078481;
RA Prade R.A., Timberlake W.E.;
RT "The penicillium chrysogenum and Aspergillus nidulans *weta*
developmental regulatory genes are functionally equivalent."
RL Mol. Gen. Genet. 244:539-547(1994).
CC -1- FUNCTION: RESPONSIBLE FOR ACTIVATING A SET OF GENES WHOSE PRODUCTS
CC MAKE UP THE FINAL TWO CONIDIAL WALL LAYERS OR DIRECT THEIR
CC ASSEMBLY AND THOUGH THIS ACTIVITY IS RESPONSIBLE FOR ACQUISITION
CC OF SPORE DORMANCY, REGULATES THE EXPRESSION OF SPORE-SPECIFIC
CC GENES.
CC -1- DOMAIN: HAS AN ACIDIC N-TERMINUS (AA 1-50) FOLLOWED BY A SER-,
CC THR-, PRO-RICH DOMAIN (AA 118-218) AND A BASIC C-TERMINUS (AA
CC 428-499).
CC -1- SIMILARITY: TO E.NIDULANS META.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X80058; CAA56364.1;
DR Developmental protein; Conidiation; Transcription regulation;
KM Activator.
FT DOMAIN 262 265 POLY-SER.
FT SEQUENCE 499 AA; 54899 MW; 0CCE6F4E15248107 CRC64;
Query Match 9.5%; Score 132; DB 1; Length 499;
Best Local Similarity 24.4%; Pred. No. 0.18;
Matches 60; Conservative 40; Mismatches 92; Indels 54; Gaps 14;
OY 50 NSSSDTYTSEDOEKGKEKDTAFQTSFDRNFEDLNSI---DIOQTI---QHQQOOP 101
DB 8 HSFNLFQNYVMKETSANDKGSAL-SDFDQFLPPLDLSLSDGDLPPVYSTKRQSQFP 66
OY 102 QOQOOLQSDTDNNLIDFSEFQTPMTSTLDTLKONPTVDKYNH-----APY---INTSP 153
DB 67 WSEWESLDDCGAADHFAFD-----TVHPSAISDVNLNFPVSPPTASHGLSTSP 118
OY 154 NKSIMKATPKKASPKKVAFTYNPEIHYPDNKRVEEEOOQKEDSVPEPLLO----- 206
DB 119 S---TPPATPKRKPTQSAI-TPKSIH-----RSPNERSHLRKQSFSLRSSNLASKA 170
OY 207 ---HOKKDPQSFVNSPEDITNASVPTPP-----LHTTKPTFAOLLKNNNEVNSEPEA 255
DB 171 RMAYPEAQAQRQ-NFSLHGSSEDRPLSPSPSDVLIQHNMPT-EQINQHOD-SAERAS 227
OY 256 LTDMKL 261
DB 228 QYDARL 233
RESULT 5
ZMSL_YEAST
ID ZMSL_YEAST STANDARD: PRT: 1380 AA.
AC P46974;

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Zinc finger protein ZMS1.
GN ZMS1 OR YJRI27C OR J2052.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=RC11-6A.
RA Thomas D., Bardey R., Surdin-Kerjan Y.;
RW Submitted (xxx-1993) to the EMBL/GenBank/DBJ databases.
RL (2)
RP SEQUENCE FROM N.A.
RA Rose M., Koeltner P., Entian K.D.;
RW Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RL (1)
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: STRONG, TO YEAST YML081W.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L26506; AAA35240.1; -
CC EMBL: 249627; CA89658.1; -
CC HSSP: P07248; 1ARD.
CC TRANSFAC: T04595; -
CC SGD: S0003888; ZMS1.
CC InterPro: IPR000822; znf_C2H2.
CC Pfam: PF00096; zf-C2H2; 2.
CC PRINTS: PR00048; ZINCINGER.
CC SMART: SM00355; znf_C2H2; 2.
CC PROSITE: PS00028; ZINC-FINGER_C2H2_1; 2.
CC PROSITE: PS00157; ZINC-FINGER_C2H2_2; 2.
CC DNA-binding: Nuclear protein; zinc-finger; Metal-binding.
FT DOMAIN 64 67
FT ZN_FING 151 173
FT ZN_FING 179 202
FT DOMAIN 424 433
FT CONFLICT 1 17
FT CONFLICT 1117 1118
FT CONFLICT 1131 1131
FT CONFLICT 1143 1380
SQ SEQUENCE 1380 AA: 155062 MW: 155062 MW: FBE987EAB5B639000 CRC64;
(IN REF. 1).
Query Match 9.4%; Score 130.5; DB 1; Length 1380;
Best local Similarity 18.9%; Pred. No. 0.7;
Matches 66; Conservative 53; Mismatches 121; Indels 109; Gaps 13;
QY 3 STPSKLLPDKSHLQLOPOSSASIFNSPTKPL-NPRTNKSRLDPNSSSDTYTSEOD 61
DB 331 NTPSSMHTKRHASFSSAMTYMSSNSPHHSITNFEVEDAPH-OVGRSTPOMTAKOL 389
QY 62 QEGKGE-----EKKDTAFQTSFDKN-FDLNSIDIOQTQHQQQPOQQQQLSOTDN 112
DB 390 MESVSELDLPRLTDEPQAIK--FNLTLPNDPS-----GQQQOQQOQQOQNSTSS 438
QY 113 NLIDEFQTPMTSLD/LTQNPVTDK-----VNEHAP----- 146
DB 439 TIYNSNNGSVATIPGVYLLSSGSPSLD/LTWNNAHAGAGGYMSSHSHSPDLGCFSHDKPT 498

QY 147 -----TYINTSP-----NKSIMKKATPKASBKVAFTVTN 176
DB 499 VSEFNLPSSPNTIPSNSTTASNSYNLANQTYROMNEQPLMSLSPKNPPTVSDSSST 558
QY 177 PEIHHPDNKRKEEDSOQOKEDSVEPLIOHOW-----KDPQSO-----FN 216
DB 559 INFNGTNNLLEPSMEPNDSNIDPAIDKMLSEFINNSDPSTEKINFENHPNDIGFI 618
QY 217 YSDEDTNAVSP-PPPLHTT-----KPPPAQLLNKNEVNSEPE 254
DB 619 YSPSSRSSIIPKSPRNHSATSLNHEKASLSPRLNLSNGSTDLSPITQ 667
RESULT 6
POM1_SCHPO
ID POM1_SCHPO STANDARD; PRT; 1087 AA.
AC 009690;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative dual specificity protein kinase pom1 (EC 2.7.1.-).
GN POM1 OR SPAC2F7.03C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean C.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voickert G., Aert R., Roden B., Gymnopoulos B.,
RA Welteens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moser D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Beer P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carulli L., Lowe T., McCombie W.R., Paulsen O., Potashkin J.,
RA Shipakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
RN (2)
RP CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=98241493; PubMed=9573052;
RA Baehler J., Pringle J.R.;
RT "Pom1, a fission yeast protein kinase that provides positional
RT information for both polarized growth and cytokinesis."
RL Genes Dev. 12:1356-1370(1998).
CC -1- FUNCTION: INVOLVED IN LOCALIZATION OF POLARIZED GROWTH AND
CC CYTOKINESIS. MAY INTERACT WITH BOTH THE ACTIN AND MICROTUBULE
CC CYTOSKELETON. REQUIRES TEAL FOR LOCALIZATION TO THE CELL ENDS BUT
CC NOT TO THE CELL CENTER.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MNB/DYRK SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Z50142; CA90490.1; -
 DR HSSP: P24941; 1CKP.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR Pfam: PF00069; pkinase.1.
 DR ProDom: PD000001; Euk_pkinase.1.
 DR SMART: SM00220; S_TKc.1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
 DR PROSITE: PS00108; PROTEIN_KINASE_DOM.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
 DR Transferase: Serine/threonine-protein kinase; ATP-binding.
 KW DOMAIN 699 995
 FT NP_BIND 705 713 ATP (BY SIMILARITY).
 FT BINDING 728 728 ATP (BY SIMILARITY).
 FT AC_SITE 825 825 BY SIMILARITY.
 FT DOMAIN 62 65 POLY-SER.
 FT SEQUENCE 1087 AA; 121146 MW; 3277EC/CA840F8BE2 CRC64;

Query Match 9.4%; Score 130; DB 1; Length 1087;
 Best Local Similarity 21.9%; Pred. No. 0.58;
 Matches 74; Conservative 34; Mismatches 108; Indels 122; Gaps 14;

QY 13 KSHLQLOPOSSASIFNS-----PTKPLNF-----PRTNSPLDPSNDDTYTS 58
 DB 247 RHTNHSSTHPRASIGKSGIPVPTIPNIGHSTHDPKANSIGSLTKSS----- 300
 QY 59 EODEKEEKKDPAFQTSFDRNPLDNDSDIQOTIQHQOQPOQOQOQSLQTDNN----- 113
 DB 301 -----ESKNLSLTQSPKLTSS-----NSF-----FKELSPHSQITLSNKNHSHVG 341
 QY 114 -LIDFSPQTPMTSTLDTLTKQNPYDKVNEHAPTYIMTSPKSIIMKATPKASPKYAF 172
 DB 342 SOTKSHSATP--SYFD--NNKPYSSDNHNNTTTSSQVHDS--RNPDPAAKRAVS- 392
 QY 173 TVTNPETHYPD--NNEEEDSOQR----- 196
 DB 393 OKTVNDGHRNHEAKHGNVQNESKQSKSSNRGGRGFFSRLSFRSSRMKKGSKAK 452
 QY 197 -EDSVPEPLIOH-----QMKDPQSFNTSDEDT--NA 224
 DB 453 HEDADVAIPAIPHAYIADSSTKSSYRNKKTPRTKSRMQQFINMKPKSKERSNSNGNSDSA 512
 QY 225 SVPTPLHTTKPTFAQLLNKNEVSEPEALTDKMLK 262
 DB 513 SPPVPRILSTRSQVSRPEKPEELIPVPLPSNFKDK 550

RESULT 7
 ID N124.SCHPO STANDARD; PRT: 1159 AA.
 AC 009904;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Nucleoporin nup124 (Nuclear pore protein nup124).
 GN NUP124 OR SPAC30D11.04C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=99340273; PubMed=10409764;
 RA Balasundaram D., Benedik M.J., Morfnew M., Dang V.-D., Levin H.L.;
 RT "Nup124p is a nuclear pore factor of Schizosaccharomyces pombe that

RT is important for nuclear import and activity of retrotransposon Tf1.";
 RL Mol. Cell. Biol. 19:5768-5784(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
 RA Brooks K., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jajels K.,
 RA James K., Jones I., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule L., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motter S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Shpakovski G.V., Usersty D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RA Nature 415:871-880(2002).

CC - FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
 CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
 CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT. IN S.POMBE IT IS
 CC REQUIRED FOR THE NUCLEAR LOCALIZATION OF RETROTRANSPOSON TFI.
 CC - SUBCELLULAR LOCATION: Nuclear pore complex.
 CC - DOMAIN: CONTAINS F-X-F-G REPEATS.
 CC - SIMILARITY: SOME TO YEAST NUCLEOPORINS NUP1 AND NUP2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AF280406; AAF90179.1; -
 DR EMBL: Z67961; CA91890.1; -
 KW Nuclear protein; Transport; Repeat.
 FT DOMAIN 270 274 POLY-ARG.
 FT SEQUENCE 1159 AA; 123974 MW; 852F84BA2E0619A3 CRC64;

Query Match 9.3%; Score 128.5; DB 1; Length 1159;
 Best Local Similarity 20.6%; Pred. No. 0.76;
 Matches 75; Conservative 50; Mismatches 128; Indels 111; Gaps 14;

QY 2 NSTPSKLPIDKXSHLQLOPOSSASIFNSPTKPLNFPTNSPLDPSNDDTYTSED 61
 DB 198 NPTCEKRKRSRSPSLSKKSVARASENPSAKQ-----KSPSGNDSHKSLDID 250
 QY 62 QEKKEE-----KKDTAFQTSFDRNPLDNDSD 87
 DB 251 KENGFEVSAAKHHVPHRSSRRRRHQRLPIIYETLEQMDLRPLVNAEYQTDSPNGMT 310
 QY 88 IDI-OOTIOHQOQPOOQOQO-----LSOTDNILDFSPQTPMTSTLDTKQNP 136
 DB 311 MFTDKODIYHRLSTPSRKROLEKCHIKAFSAVDEDLDEIFACEDVHYTA-LPKQNK 369
 QY 137 VDKVENEHAPTYINTSPKSIIMKATP-----RASPKKVAFTYNPE----- 178
 DB 370 SRLIEPIIASPKDNTSDKGLTKSAPTFELOASTIPKPVKTSPTDVALTIANADNKT 429

```

OY      179   IHNVDPNRVEEDSO-----OKDVSPEPLIDHOMKDEPOFVNSD-----    219
Db      430   EHHOPUSKTEPAKPSQFSSPTKESTRTSEVEPSPSKELKS-SHEFSPEFKEPKTEA 488
OY      220   -BDTNASVP-----TPPL-----HTTTPPT-FAQLLNKNNEVNS-----EPEALIDM 259
Db      489   TUDKLNVPKFEFKPTATADVOGTNRLKEMERKPTEFFAQAPSKTGTSPTLENKKPFESFSL 548
OY      260   KLRK 263
Db      549   SPKR 552

RESULT 8
YJ53_YEAST
ID      YJ53_YEAST          STANDARD;          PRT;          309 AA.
AC      P47129;
DT      01-FEB-1996 (Rel. 33, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      Hypothetical 35.5 kDa protein in MIRI-STB18 Intergenic region.
GN      YJR083C OR J1857.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288c;
RX      MEDLINE=96437976; PubMed=8840504;
RA      Huang M.-E., Manus V., Chaut J.-C., Galibert F.;
RT      "Analysis of a 62 kb DNA sequence of chromosome X reveals 36 open
RT      reading frames and a gene cluster with a counterpart on chromosome
RT      XI.";
RL      Yeast 12:869-875(1996).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: Z49583; GAB89610.1; -.
DR      EMBL: L47993; AAB39306.1; -.
DR      SGD: S0003843; YJR083C.
DR      Hypothetical protein.
SEQUENCE 309 AA; 35520 MW; FE29EBD0192B2113 CRC64;

Query Match          9.0%; Score 125; DB 1; Length 309;
Best Local Similarity 22.5%; Pred. No.0.29;
Matches 64; Conservative 36; Mismatches 124; Indels 60; Gaps 10;

OY      7     KLLEPIDKSHLD-----LQPSSASIFNPTRKPLNPRPNNSKPSLDPNSSSPTY 56
Db      16     KLSIVDKHSOGQQQRPHQKHENVEPESKPRV----TTPLK-PKLAIPISSPGQSTNQ 70
OY      57     TSEDQOEKKEKKDTAFOTSFDRNFDDLSIDIOQTIOHQOOQPOQOOQLSOTDNMLID 116
Db      71     SPVSDHASPISTDODLIYLAKHREINELSPKLEVA-----OKELKOLBL---- 116
OY      117     EPSFGTPMTSIDTLKONFT--VDKVNEHNATY--INTSPKSIIMTKATPKAAASKYAF 172
Db      117     QKKDPLPRNGOOKLGNNSEYLSFTFKTIQTFADVNNSPMLKGKKSINDFSK---- 172
OY      173     TVTNEEIHYPNRVEEEDSOOKEDSVERPILQHQMKDPSOFNYSDEBDTNASVP-TTP 231
Db      173     -----PNNVNNSNTNNTLPRPKRPDRPPNRGQRMNINAPRSSSEBTTPISGGPLLP 222
OY      232     LHFTTK-----PTFAQLLNKKNNEVNSEPALIDMKLR 263

```

```

DB 223 RNTMKNANTTATAGENTPELQRIILINKFNQNMNDEDEFDLLER 266

RESULT 9
ID HMW1_MYCGE STANDARD: PRT: 1139 AA.
AC 049413; 049365;
DT 01-NOV-1997 (Rel. 35, created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Cytoadherence high molecular weight protein 1 (Cytoadherence accessory
DE protein 1).
DE GN HMW1 OR MG312.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischman R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky G., Fuhmann J.L.,
RA Tomb J.-F., Dougherty B.A., Boff K.F., Hu P.-C., Lucher T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 721-847 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Boff K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
RN [3]
RP -I- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
RC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
RX CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
RC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERESIN PROTEINS
RC IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
RC SIMILARITY).
RN [4]
RP -I- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
RC MEMBRANE (BY SIMILARITY).
RN [5]
RP -----
RC This SWISS-PROT entry is copyright. It is produced through a collaboration
RC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RC the European Bioinformatics Institute. There are no restrictions on
RC use by non-profit institutions as long as its content is in no way
RC modified and this statement is not removed. Usage by and for commercial
RC entities requires a license agreement (see http://www.isb.slb.ch/announce/).
RC or send an email to license@isb.slb.ch.
RN [6]
RP EMBL: U39712; AAC71534.1; -
DR EMBL: U02261; AAD12527.1; -
DR TIGR: MG312; -
RN [7]
RP Cytoadherence: Structural protein: Complete proteome.
SQ SEQUENCE 1139 AA; 130531 MW; 0011D3288C3D856 CRC64;

Query Match 8.9%; Score 123; DB 1; Length 1139;
Best Local Similarity 24.6%; Pref. No. 1.6;
Matches 66; Conservative 39; Mismatches 103; Indels 60; Gaps 15;

QY 11 IDKHS---HQLQPOSSASINSPKPLNFPRTNSKPSLDPNSSD--TYSSEDQDEKG 65
DB 228 WQPSDDHFAQPEST-----TDSYSPDSJDRPFLDPSLDHNVQNMNDEHKEK 279
QY 66 KEKKDPAFQTSFDR-NEDLNSIDIOQTIGNQOQOPOQOOLSTDNMLDEFSFGTQM 124
DB 280 PVAEDSNKQVGFDOVQNLNDEEQLPAEKVTTTDESKAQQVYD-----SYLPI 332
QY 125 -TSTLDT-----KQNPV---DKVYENHADPTJINTSPKSIKMKAPKASPK- --VA 171

```

Db 333 DTDDQDTFFSSFTQPTVEFDQVNSE-----VNDQFKPEITKEPVLESSFKNQDYVE 387
QY 172 FTVTNPETIHHPDNRVEEEDSOQKEDSEVPEPLIQHOMKDPDSQFVSDDEDTASV---PP 228
Db 388 TSDLNSESNLXSEN-----NKDATNNDSDLNSEFT-----QLNSNETASDVHYEAK 434
QY 229 TPLPHTTKPTFAQLL---NKNNEVNSEP 253
Db 435 SEPIHDYK--FGSDLSQSNNSNLSLESEP 460
RESULT 10
P200_MYCGE STANDARD: PRT: 1616 AA.
ID 049429: 049298: 049352: 049353;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein P200.
GN MG386.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann J.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
RA "The minimal gene complement of Mycoplasma genitalium.",
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 256-427; 432-543 AND 1083-1140 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RA "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.",
RL J. Bacteriol. 175:7918-7930(1993).
CC -I- FUNCTION: COULD BE A ACCESSORY STRUCTURAL COMPONENT IN
CC CYTADHERENCE (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: U02245: AAA03400.1; -;
CC DR EMBL: U02245: AAA03400.1; -;
CC DR EMBL: U02175: AAD12458.1; -;
CC DR EMBL: U02126: AAD12402.1; -;
CC DR TIGR: MG386; -;
CC KM Cytadherence; Structural protein; Repeat; Complete proteome.
FT DOMAIN 1205 1389 2 X 32 AA REPEAT.
FT REPEAT 1205 1236 1-1.
FT REPEAT 1358 1389 1-2.
FT DOMAIN 891 1389 2 X 26 AA REPEAT.
FT REPEAT 1161 1186 2-1.
FT REPEAT 1310 1339 2-2.
FT REPEAT 256 256 P -> S (IN REF. 2).
FT CONFLICT 304 304 S -> F (IN REF. 2).
SQ SEQUENCE 1616 AA; 185678 MW; 6AF76A13AC49E4FF CRC64;

Query Match 8.9%; Score 123; DB 1; Length 1616;
Best Local Similarity 20.4%; Pred. No. 2.4;

Matches 49; Conservative 52; Mismatches 111; Indels 28; Gaps 9;
QY 4 TPXKLLPIDKHSL---QIQP-QSSASITFNSPTKPLNPRNRSKSL-----DP 49
Db 1202 TQPEIQPVESQPEAFPTVQPTQPEAKFSDVEVEQPESSSEPTQOHVSEASFDP 1261
QY 50 NSSDPYTSSEQDQEKKEKKTAFQTSFDR-NFIDLN-SIDIQTIHQOQOQOQOOL 107
Db 1262 NTFDEPNNDPQOPSDSLQSPSPQYDVERNYDPDERNYLEKSPSEPEPEVQEQP 1321
QY 108 SQTNNLIDFSPQTPMTSTLDLTKONPTVDKVENNHADPTINTSPKSIKKAIPKASP 167
Db 1322 GAVEPESAFAKFDSEVESVD-SQEPLEVEQTPQPEIQPVESQPE-EATFTVQPEQTP 1379
QY 168 KQVAF-----TYTNPEIHHPDNRVEEEDSOQKEDSEVPEPLIQHOMKDPDSQFVSDDEPT 222
Db 1380 QEAKFDSFVETIQPEQVSSSEPEVVOQPNFEKRPETVLEP-----QADEIQPEASEES 1434

RESULT 11
YHCB_YEAST STANDARD: PRT: 605 AA.
ID YHCB_YEAST
AC P38739;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Hypothetical 63.8 kDa protein in GUT1-RIM1 intergenic region
DE precursor.
GN YH1028W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Doyer J.,
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaubin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.",
RL Science 265:2077-2082(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: U11583: AAB65040.1; -;
CC DR EMBL: S48940: S48940.
CC DR SGI: S0001020: WSC4.
CC DR InterPro: IPR002889; WSC.
CC DR Pfam: PF01822; WSC.1.
CC DR SMART: SM00321; WSC.1.
CC KM Hypothetical protein; Transmembrane; Signal.
FT SIGNAL 1 26
FT CHAIN 27 605 POTENTIAL.
FT DOMAIN 116 317 HYPOTHELTICAL PROTEIN YH1028W.
FT TRANSMEM 415 435 SER/THR-RICH.
FT POTENTIAL
SQ SEQUENCE 605 AA; 63807 MW; C9DF8C8AA9553811 CRC64;

Query Match 8.7%; Score 121; DB 1; Length 605;
Best Local Similarity 22.9%; Pred. No. 1.1;
Matches 65; Conservative 32; Mismatches 85; Indels 102; Gaps 10;

Db 703 SYTNPSIQSPNLSFQMANXYFQSL 729
Search completed: March 17, 2003, 12:25:03
Job time : 13.2756 secs

```
RC STRAIN-S288c:
RA Wang W., Zheng L., Chan C.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-GRF88;
RA Song J.M., Cheung E., Rabinowitz J.C.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA MEDLINE-97051586; Pubmed-8896263;
RA Parle-Mcdermott A.G., Hand N.J., Goulding S.G., Wolfe K.H.;
RT "Sequence of 29 kb around the PDRI1 locus on the right arm of
RT Saccharomyces cerevisiae chromosome XV: similarity to part of
RT chromosome I."
RL Yeast 12:999-1004(1996).
CC -1- FUNCTION: Involved in vesicular transport at a late stage of the
CC secretory pathway
CC -1- SUBUNIT: POTATIVE 10-FORMYL-TETRAHYDROFOLATE BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL: U03492; AAB09719.1; -
DR EMBL: UA2227; AA85443.1; -
DR EMBL: Z49821; CA88976.1; -
DR EMBL: Z75237; CA99650.1; -
DR SGD: S0005856; SCD5.
KW Transport; Protein transport; Membrane; Repeat.
FT DOMAIN 405 448 3 X 20 AA APPROXIMATE REPEATS.
FT REPEAT 405 424 1-1.
FT REPEAT 439 458 1-2.
FT REPEAT 479 498 1-3.
FT DOMAIN 534 728 9 X 12 AA APPROXIMATE REPEATS.
FT REPEAT 534 545 2-1.
FT REPEAT 564 575 2-2.
FT REPEAT 593 604 2-3.
FT REPEAT 608 619 2-4.
FT REPEAT 623 634 2-5.
FT REPEAT 636 647 2-5.
FT REPEAT 650 661 2-7.
FT REPEAT 683 694 2-8.
FT REPEAT 717 728 2-9.
SEQUENCE 872 AA; 97305 MW; C60F5BE808E1D31 CRC64;

Query Match 8.6%; Score 119.5; DB 1; Length 872;
Best Local Similarity 22.5%; Pred. No. 2;
Matches 60; Conservative 39; Mismatches 115; Indels 53; Gaps 13;

QY 4 TPESKLPIDKSHQL-----QOQSSASIFNSPTKPLNPRNRSKPSLDPNSSSDITYS 58
DB 502 SPVSNNPVSSMFOAQFTNOSSSQSTGPAFLNSPNTL-----PQSNQOOPYQ 549
QY 59 EDOEKGEKEKKDTAFQTSFDRNFDLNN-----SIDIOQTIOHQOQOPQOQOQLSOTD 111
DB 550 EVNPTQAKIPSPNISPOHTYSNNVRINGNIVSPKVEITGAPPPQNTLPQHQOQSHLSP 609
QY 112 NNLIDEFSTQTPMTSLDLTKONPTVDKVENHNAPTYINTSPNKSINKKATPKASPKVA 171
DB 610 QNTIPQHQ-RSOLISPONTFTQNPQI--LSPQH--TVSN--NQATM-----ISPQN-T 654
QY 172 FTVTNPEIHHYPPDNRYEEDQSOQKEDSVPEPLIOHQWKDPSQFNSDEDTNASVPTPP 231
DB 655 YTNNOQOPQHLPPEPPPPRAOQOQO--GAIVPP-QHM-----YSNVOKNNLVPTOP 702
QY 232 LHTTKPTF--AQLLNKNNENVSPEAL 256
```


Db 368 TSDLNSESILYSEN-----NKDATNNDSINSEFI-----QLNSNETASDVHYESK 434

Qy 229 TPELHTTKPTFAQL--NKNEVNSEP 253

Db 435 SEPIHDYK--FGSDLSQSNNSNLSSESP 460

RESULT 2

US-09-820-843A-16

Sequence 16, Application US/09820843A

Publication No. US20030039963A1

GENERAL INFORMATION:

APPLICANT: Council of Scientific and Industrial Research

TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN

FILE REFERENCE: 063915

CURRENT APPLICATION NUMBER: US/09/820,843A

CURRENT FILING DATE: 2001-03-30

NUMBER OF SEQ ID NOS: 118

SOFTWARE: Patentin version 3.0

SEQ ID NO 16

LENGTH: 1616

TYPE: PRT

ORGANISM: M. genitalium

FEATURE:

NAME/KEY: misc.feature

OTHER INFORMATION: cytochrome-accessory protein

NAME/KEY: misc.feature

OTHER INFORMATION: g111046097

US-09-820-843A-16

Query Match 8.9%; Score 123; DB 9; Length 1616;

Best Local Similarity 20.4%; Pred. No. 0.71;

Matches 49; Conservative 52; Mismatches 111; Indels 28; Gaps 9;

Qy 4 TPKSLPIDKHSHL--QLOP-QSSASIFNSPTKPLNPRTPNSKPSL-----DP 49

Db 1202 TPELHTTKPTFAQL--NKNEVNSEP 253

Qy 50 NSSDPTYSSEDOEGKEKKKDTAFQTSFDR-NFDLDN-SIDIOQTIOHQOQOPOQOQOL 107

Db 1262 NYDFEPNVDFOQPSYSDLOQPSEFOYDVEDENYDFEDENYDESPSEPEQVEQOP 1321

Qy 108 SOTDNLLIDFSEOTPMSTLDLTKQNPYVDKVNENHATYINTSPKNSIMKATPKKASP 167

Db 1322 GAVERPSEAKFDSVPSVOD-SQPEPLEVQTOPEIQVESQP-ETFPDVOPEQTP 1379

Qy 168 KKAFA-----TYTNEIHHYPPNRYEEDOSQOKEDSVPEPLIQHOMKPSOFNSDEDT 222

Db 1380 QEAKFDSVETIQEFOVSEPEVYVQPNFEKRPETVLEP-----QADEIQPEASEES 1434

RESULT 3

US-09-854-133-425

Sequence 425, Application US/09854133

Publication No. US20020183499A1

GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.

APPLICANT: Mohamath, Radooh

APPLICANT: Henderson, Robert A.

APPLICANT: Benson, Darin R.

APPLICANT: Secrist, Heather

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.475C10

CURRENT APPLICATION NUMBER: US/09/854,133

CURRENT FILING DATE: 2001-05-11

NUMBER OF SEQ ID NOS: 735

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 425

LENGTH: 4019

TYPE: PRT

ORGANISM: Homo sapiens

US-09-854-133-425

Query Match 8.9%; Score 123; DB 9; Length 4019;

Best Local Similarity 19.9%; Pred. No. 2.1;

Matches 70; Conservative 40; Mismatches 116; Indels 126; Gaps 15;

Qy 2 NSTPSKLPIDKHSHLOLOPOSSASIFNSPTK---PLNFPRTNSKPSLDP----- 49

Db 2393 NSAP-----AHLPLPPIPIQPIAOLPIKTCTPAGTVSNANPOSGPPRVEFDON 2443

Qy 50 NSSDPTYSSEDOEGKEKKKDTAFQ--TSFDRN----- 81

Db 2444 NPSEFQERERKRRERQOERIQLOMVEDRORALQRMEMEDHGWSISSSRTSV 2503

Qy 82 -----FDLNSIDIOQTIOHQOQOPOQOQOLSQT-DNNLIDFSEFOPTMTSLDLTKQ-- 133

Db 2504 SQIPFSSDLPDQFQPLRLOOSPOHQOQOVLQOQOIQOQSINSPTQFMOTNEHR 2563

Qy 134 -----NPTVKNENH---APTYINTSPKNSIMKATPKKASPKKVAFT 173

Db 2564 QVGPSEFVPSIPVSGNFSVKGHGLSGTSFOQSPVPSFTPALPAPP----- 2617

Qy 174 VTNPET-----HHYPNDR-----VEEE-----DOSOKEDSVPEPLIQ 206

Db 2618 VANSILPCGQDSTITRHASYPOSTOSLIDLYSDIIPEEKGKKRTRKKRRDDAE----- 2672

Qy 207 HQMKDPSOFNYSDEDTNANVPPTPLHTTKPTFA-----QLKNNEVNSEP 253

Db 2673 -STKAP-----STPHSDIAPPTPGISETTSTPAVSTPSELQADQESVEP 2718

RESULT 4

US-09-738-973-425

Sequence 425, Application US/09738973

Patent No. US20020110563A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Henderson, Robert A.

APPLICANT: Lodes, Michael J.

APPLICANT: Fling, Steven P.

APPLICANT: Mohamath, Radooh

APPLICANT: Algate, Paul A.

APPLICANT: Secrist, Heather

APPLICANT: Indrias, Carol Joseph

APPLICANT: Benson, Darin R.

APPLICANT: Elliott, Mark

APPLICANT: Mannion, Jane

APPLICANT: Kalos, Michael D.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.475C9

CURRENT APPLICATION NUMBER: US/09/738,973

CURRENT FILING DATE: 2000-12-14

NUMBER OF SEQ ID NOS: 587

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 425

LENGTH: 4019

TYPE: PRT

ORGANISM: Homo sapiens

US-09-738-973-425

Query Match 8.9%; Score 123; DB 10; Length 4019;

Best Local Similarity 19.9%; Pred. No. 2.1;

Matches 70; Conservative 40; Mismatches 116; Indels 126; Gaps 15;

Qy 2 NSTPSKLPIDKHSHLOLOPOSSASIFNSPTK---PLNFPRTNSKPSLDP----- 49

Db 2393 NSAP-----AHLPLPPIPIQPIAOLPIKTCTPAGTVSNANPOSGPPRVEFDON 2443

Qy 50 NSSDPTYSSEDOEGKEKKKDTAFQ--TSFDRN----- 81

Db 2444 NPSEFQERERKRRERQOERIQLOMVEDRORALQRMEMEDHGWSISSSRTSV 2503

[illegible]

```

RESULT 5
US-09-801-368-428
: Sequence 428, Application US/09801368
: Patent No. US20020128250A1
: GENERAL INFORMATION:
: APPLICANT: Busby, Robert
: APPLICANT: Cali, Brian
: APPLICANT: Hecht, Peter
: APPLICANT: Holtzman, Doug
: APPLICANT: Madden, Kevin
: APPLICANT: Maxon, Mary
: APPLICANT: Milne, Todd
: APPLICANT: No. US20020128250A1man, Thea
: APPLICANT: Royer, John
: APPLICANT: Salama, Sofie
: APPLICANT: Sherman, Amir
: APPLICANT: Silva, Jeff
: APPLICANT: Summers, Eric
: TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
: FILE REFERENCE: 109272.147
: CURRENT APPLICATION NUMBER: US/09/801,368
: PRIOR APPLICATION NUMBER: US 09/487,558
: PRIOR FILING DATE: 2000-01-19
: PRIOR APPLICATION NUMBER: US 60/160,587
: PRIOR FILING DATE: 1999-10-20
: NUMBER OF SEQ ID NOS: 440
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 428
: LENGTH: 605
: TYPE: prt
: ORGANISM: Saccharomyces cerevisiae
US-09-801-368-428

```

```
Query Match          8.7% Score 121 DB 10 Length 605;
Best Local Similarity 22.9%, Pred. No. 0.31;
Matches    65; Conservative   32; Mismatches    85; Indels   102; Gaps     10;
```

OY 21 POSSASAFNSPTKPLNPRNTSKPSLDPNSSDDVYTSEODKEGKEKKDPATQSPDR 80
 | ||| : | : | : | : | : | : |
Db 268 PTSSASPRTSSMTPTSTFTTTSTPSTAPSSVTWTYS-----TTSNP 310

OY 81 NFEDLNSIDIQOTIGHOQQOQPQOOQQLSCTDNLLIDESFQTPMSTLDLTQNPTVDKY 140
 | : : : : | : : : : | : : : : | : : : : |
Db 311 ITSTITSVNLQTSLEY-----SVITTVSVHMPDINISEITSTRYLTMKVY 354

OY 141 -----NNHAPTYINTSPNKSIIMKAIRPKASKPVY 170

Db 335 ITOIISLTGLATPTSAVATTSAVSOGRIITANNNSITTNSMTPTNSTEKERGY-WDSPGT 413

OY 171 AFT-----VTNPELIHHYPDMRVEEEDSOQOKEDSVERPLLIOHQMKDESFNY 217
 | | : | : | : | : | : | : | : | : |
Db 414 ATFVVGVGCVLIIICILDIYLHNH--RTRARRAKODEENELYQSKFYISK-----Y 462

OY 218 SDEDTNASYVPPTPLHTTKPTFFAOLLKNNEVNSEADAL-TDMK 260
 : | : | : | : | : | : | : | : | : | : | :

Db 463 PNEVT-----TTTTLHTPPS-----SNSTFSTPRLTYTDEK 493

RESULT 6
US-09-815-242-11497
; Sequence 11497, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION

```

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELTRA.011A
CURRENT FILING DATE: 2001-03-21
CURRENT FILING DATE: 2001-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-05-23
PRIORITY FILING DATE: 2000-05-23
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: 2000-10-23
PRIORITY FILING DATE: 2000-11-27
PRIORITY FILING DATE: 2000-12-22
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11497
LENGTH: 844
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-815-242-11497
```

Query Match	8.7%;	Score 120;	DB 10;	Length 844;
Best Local Similarity	20.7%;	Pred. No. 0.53;		

Matches 51; Conservative 42; Mismatches 117; Indels 36; Gaps 8;

```
Qy 28 IFNSPTKPLNFPRNTSKPSL-----DPNSS-----SDTYTSEDQEKCKEEK 69
      : | : : | | | | : : | : : :
Db 134 LFKLPKSVYPPYMNKTQSLKLEIYKQCLQAFSPNFSLKGEFENTPDSQCKKETNNDKE 193
```

```
QY 70 KDTAFQTSFDRNFOLDNS---IDIQQTIQHQQQPQQQQQLSQTDNNLIDEFSEQTPMTS 126
| : : | | : | : | : | : :
Db 194 KENLKBNPIDENHNTPNESEFLAIPTPYNTLLNNESEPOEGVOISPHBPTHYTYIPKRN 253
```

Qy 127 TLDLTQNPYVDKVNENHAPYYINTSPKNSIMKATPKASPKKVAFTYNPEIHHYPDR 186

QY 187 VEEEDSQOKEDSVEPPLIQHOKWDPQSFNYSDSDJTNASVPTPTPLHIT--KPTFAQLL 243

DB	304	-HKIPNPIKEDDQESPOENPOKENIBENIEERETIQ-NAPSFSPLTISAKKPYMKEL	361
OY	244	NKNEV	249

Db 362 SENKEI 367

RESULT 7
US-09-765-272-218
; Sequence 218, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:

```

; APPLICANT: Chol et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 218:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 218:
US-09-765-272-218

Query Match      8.1%; Score 112.5; DB 10; Length 565;
Best Local Similarity 22.0%; Pred. No. 1.1;
Matches 61; Conservative 37; Mismatches 102; Indels 77; Gaps 13:

QY 1 MNSTPSKLLPIDKSHSLQLOQSSA--STFNSTPKPLNFPRTNSKPSLDPNSSSDTYTS 58
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 78 VESKPEEKVAVKPE-----QPSDKPAESKVEQAGEPV-APREDKAPVEP----- 123
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 59 EQDQKGEKKDPAFQTSFDRNFDLNSIDIQOTIQHQQQPQQ-----QQLSQTQNN 113
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 124 EKOPAPEEKA-----VEETPKQESTPDTKAEETVPEKKEEVNQ 164
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 114 LIDERSFQTPMTSLDLTKONPTVDKVNENHAPTYINTSPNKSIMKKA7PKAPKVAFT 173
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 165 SIEQPKVTPAVEKQETPEEKVQAGEPAVAPREDEQAP-----TAPVEPEK----- 212
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 VTNPETIHYPNRRVEEDQSQQKEDSV--EPPILQHQMKDSQFNYSDEDTNNAVSPPT- 229
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 213 --QPEVPE-EKAVAEETPKEDKIKIGTKEP-----VDKSELN-NQIDKASSVSPD 261
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 -----PPLHTTKPTFAQLLNKNNEVNSEPAL 256
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 262 YSTASYNALGPVLETAKGVASEPVKQPEVNSETKL 298
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
US-09-887-828A-2
; Sequence 2, Application US/09887828A
; Patent No. US20020128456A1
; GENERAL INFORMATION:
; APPLICANT: Amidon, Benjamin Stone
; APPLICANT: Bulawa, Christine Ellen
; TITLE OF INVENTION: CANDIDA ALBICANS KINASE GENES AND
; TITLE OF INVENTION: POLYPEPTIDES AND USES THEREOF
```

```

; FILE REFERENCE: 06286-291001
; CURRENT APPLICATION NUMBER: US/09/887,828A
; CURRENT FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: US 60/213,621
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 2
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-887-828A-2

Query Match      8.1%; Score 112; DB 10; Length 892;
Best Local Similarity 22.7%; Pred. No. 2.1;
Matches 57; Conservative 37; Mismatches 89; Indels 68; Gaps 11:

QY 48 DPNSSDITYTSEQDEKGEKKDPAFQTSFDRNFDLNSIDIQOTIQHQQQPQQQQQL 107
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 285 DNRSSRDSVFILNENE--QEDSVDEIVNKNLDGTV-----HNQQQHQHQQQQ 329
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 SQTD-----NWLID-ESFQTPMTSLDLTKONPTVDKVNENH----- 144
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 330 QQNDQHDHGHNNQIQVKNWEDFS-LSSREYIMENDIDLKVNQQYNNNFDTOQGNTYSTL 388
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 145 ----APTYIN-----TSPNKSIMKKA7PKAS--PKVAFTV7NPETIHYPNRRVEEDQ 192
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 389 PTLKAPPTKNNPLATNQSMTALKNNSKSTVEYKSLQML-----FEENSEQEDL 440
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 193 SQQKE-----DSVEPPLIQHQMKDPSQFNYSDEDTNNAVSP-----TPLHTTKPTFAQL 243
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 441 APPKSNVNIEMERESPTEIPDMDSLQPLPSASATNTNHPQLNKPPALYHSQASGNLE 500
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 NKNNEVSEPE 254
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 501 SRFSSLNTSPK 511
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-09-893-519A-60
; Sequence 60, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESTIVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAQ, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/IG548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Candida albicans
```

```

QY 187 VEEEDSQQRKEDSVEPPIQH---QMDPSQ-----FNFS 218
      | : | : | : | : | : | : | : | : |
Db 2608 GAETSLSDQNDKNSKYLQMAATAYNNELDQSACNPTYNSSHLFGTSTPYSAWCYYQYS 266

```

Db 2668 NSNGNAITQTYÖGITSEYEVDPSPGLLTVAŠTAÖGTHSN 2707

RESULT 11
US-09-839-996-6
; Sequence 6, Application US/09839996

```
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; Falkow, Stanley
; TIME OF INTERVIEW: November 1968
```

Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Elchr. Habbach West Albitton 6 Harbort

SYMBOL: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States

```

; ZIF: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/839,996
;

```

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791

ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249

```

; INFORMATION FOR SEQ ID NO: 6:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 1848 amino acids
;     CD: 1

```

```

;          TOPOLOGY: unknown
;          SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-839-996-6

```

Query Match	7.9%;	Score 109.5;	DB 9;	Length 1848;
Best Local Similarity	22.8%;	Pred. No. 7.5;		
Matches 57;	Conservative 34;	Mismatches 120;	Indels 39;	Gaps

```

QY 18 QLOPSSSSASIFNSPKPKPLNFBRTNSRPSLDPPSSSDIVTSSEQOEKGEKKKKDPAFQTS 77
      + + + + + : : : + + + + + : : : + + + + + : : : + + + + +
Db 1347 QAOPTOSTAV--PT--TGETAANSKPKPAKPOAQAKPOTEPARENVSIVNTKEPOSQTS 140

```

```
qy /8 FDRNF.DLNS.IDIQQQIQHQOQQPQQQQQLSQ--TDNNLLDEFSSQTPMTSTLDLTQRNPt 136
      | : | : | | : | | | | :
Db 1402 -----ATVSTEQAKEITSSNVEQAPENSINTGS-ATTMETAE-KSDKPQ 144
```

```

13/ VDAVNENHAP11INSFNSJMMNAIPASFAVAFIVINPELHIPIDNRKEE----EDQ 192
   :: | | | : | : | : | : |
Db 1446 MEVTIENDRQPEANTVADNSVANNSESESESKRRRSYQPK-----ETSAEETIVASTQ 150

```

```

150 300000SVETFLI0FM00DF0FNI3DE01NSVFF1FF011NF1FA0LUNN0N0VNS0 232
      ::||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1501 ETTVDNSVSP-----KPRSRTRRSYQTNSTPEVPELTENAE-NAENYQSGNNVANSQ 155

```


ORGANISM: Staphylococcus aureus
US-09-815-242-5816

Query Match 7.8%; Score 108; DB 10; Length 2478;
Best Local Similarity 22.0%; Pred. No. 13;
Matches 60; Conservative 34; Mismatches 95; Indels 84; Gaps 13;

31 SPTKPLNPR-----TNSKPSLDPNSSSDTYTSODODKGEKKDFAFGTS 77
DB 2112 SPTTNNNSDAATGETTATSATDANDKPOANNSSVDASTNSPTMDN----- 2158
QY 78 FDRNFDLNSIDIQOTIQHQOQOPOQOQOLSQTDNNLIDEFSGQTMTSLDLTKONPTV 137
DB 2159 -----DVTSKPEVESTNNGTDDKP-----VTETDN-----ATPAEST---TNNNSTT 2197
QY 138 DKVENHAPTINTSPNKSIMKKATPKASPKVAFVTNPELHHYPDN--RVEEDQSQOK 196
DB 2198 TATNEN-APTGSTAAPTASTEAASSADSK-----DNASVNSKQNEVNNASBSOST 2250
QY 197 EDSVEPPLIOHOKDPSOFNYS-----EDTNASVP-----PPPLHTTKPTFAQLL 243
DB 2251 NDKVAQPKSENKAK--AEKDGSDSTNQSWESTTETLPSADITEPVPNSITSKDEESTT 2308
QY 244 NK-----NNEVNSBP-EALTDMLK 262
DB 2309 NOTDAGOLKSETNVASNEADKSPSKADTEVS NK 2341

RESULT 15
US-09-815-242-12967

Sequence 12967, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes In
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12967
LENGTH: 2478
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12967

Query Match 7.8%; Score 108; DB 10; Length 2478;
Best Local Similarity 22.0%; Pred. No. 13;

Matches 60; Conservative 34; Mismatches 95; Indels 84; Gaps 13;

QY 31 SPTKPLNPR-----TNSKPSLDPNSSSDTYTSODODKGEKKDFAFGTS 77

DB 2112 SPTTNNNSDAATGETTATSATDANDKPOANNSSVDASTNSPTMDN----- 2158
QY 78 FDRNFDLNSIDIQOTIQHQOQOPOQOQOLSQTDNNLIDEFSGQTMTSLDLTKONPTV 137
DB 2159 -----DVTSKPEVESTNNGTDDKP-----VTETDN-----ATPAEST---TNNNSTT 2197
QY 138 DKVENHAPTINTSPNKSIMKKATPKASPKVAFVTNPELHHYPDN--RVEEDQSQOK 196
DB 2198 TATNEN-APTGSTAAPTASTEAASSADSK-----DNASVNSKQNEVNNASBSOST 2250
QY 197 EDSVEPPLIOHOKDPSOFNYS-----EDTNASVP-----PPPLHTTKPTFAQLL 243
DB 2251 NDKVAQPKSENKAK--AEKDGSDSTNQSWESTTETLPSADITEPVPNSITSKDEESTT 2308
QY 244 NK-----NNEVNSBP-EALTDMLK 262
DB 2309 NOTDAGOLKSETNVASNEADKSPSKADTEVS NK 2341

Search completed: March 17, 2003, 12:34:28
Job time : 14.412 secs

This Page Blank (uspto)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:20:59 ; Search time 5.1863 Seconds
(without alignments)
4875.027 Million cell updates/sec

Title: US-09-964-858-1_COPY_1_263

Perfect score: 1386

Sequence: 1 MNSPSPKLPIDKSHSLQLO.....NKNNEVNSEPALTDMLKR 263

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1386	100.0	1664	T18216	Integrin-like prot
2	141.5	10.2	236	B90577	lipoprotein b (imp
3	138.5	10.0	1166	T28680	fibronectin-binding
4	137	9.9	947	T08605	hypothetical prote
5	134	9.7	1141	E89824	hypothetical prote
6	132	9.5	499	S46660	weir protein - pen
7	130.5	9.4	1380	S57150	ZMS1 protein - yea
8	130	9.4	1087	S58147	protein kinase - f
9	128.5	9.3	1159	S62562	probable nuclear p
10	127.5	9.2	1335	T30211	autolysin E - Stap
11	127	9.2	736	T41259	hypothetical prote
12	127	9.2	1338	T18416	hypothetical prote
13	125	9.0	309	S57102	hypothetical prote
14	125	9.0	646	D82493	conserved hypothet
15	124.5	9.0	773	F90537	lipoprotein (import
16	124.5	9.0	3848	T17414	tipc protein - s
17	123.5	8.9	1094	S49313	protein kinase - s
18	123.5	8.9	1254	T18272	kinesin heavy chai
19	123	8.9	619	D90072	hypothetical prote
20	123	8.9	1139	E64234	cytadherence-acces
21	123	8.9	1616	G64242	cytadherence-acces
22	123	8.9	2541	T29340	hypothetical prote
23	121.5	8.8	529	T50609	hypothetical prote
24	121.5	8.8	948	T24445	hypothetical prote
25	121.5	8.8	6642	T29757	protein UNC-89 - C
26	121	8.7	368	G88636	protein W09612.7 l
27	121	8.7	522	T40520	hypothetical prote
28	121	8.7	605	S48940	hypothetical prote
29	120.5	8.7	910	S73361	dnad homolog prote

30	120.5	8.7	916	2	S22864	DNA topoisomerase
31	120	8.7	844	2	B71944	septum formation p
32	120	8.7	1015	2	T13062	CLOCK protein - fr
33	120	8.7	1457	2	T14577	protein kinase Yak
34	119.5	8.6	599	1	A64235	cytadherence-acces
35	119.5	8.6	872	2	S62061	SCDs protein - yea
36	119	8.6	1186	2	S72229	meiotic recombinat
37	118.5	8.5	1390	2	T14004	trfa protein - sil
38	118	8.5	411	2	T40538	hypothetical prote
39	118	8.5	462	1	A60746	chromogranin A pre
40	118	8.5	842	2	S60402	protein kinase CLA
41	118	8.5	879	2	S64755	hypothetical prote
42	118	8.5	1089	2	T14576	nosa protein - sil
43	117.5	8.5	1274	2	A89959	hypothetical prote
44	117	8.4	1062	2	G86325	hypothetical prote
45	116.5	8.4	630	2	T38023	probable transcript

ALIGNMENTS

RESULT 1

T18216

Integrin-like protein alpha chain - yeast (Candida albicans)

C:Species: Candida albicans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000

C:Accession: T18216

R:Gale, C.; Finkel, D.; Tao, N.; Meinke, M.; McClellan, M.; Olson, J.; Kendrick, K.;

Proc. Natl. Acad. Sci. U.S.A. 93, 357-361, 1996

A:Title: Cloning and expression of a gene encoding an integrin-like protein in Candida

A:Reference number: 206510; M0ID:96133936; PMID:8552638.

A:Accession: T18216

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1664 <GAL>

A:Cross-references: EMBL:U35070; NID:g1144530; PID:g1144531; PIDN:AAA96019.1

C:Genetics:

A:Gene: alpha INT1

Query Match	100.0%	Score 1386;	DB 2;	Length 1664;
Best Local Similarity	100.0%;	Pred. No. 1.1e-81;		
Matches 263;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
QY	1	MNSPSPKLPIDKSHSLQLOPOSSASIFNSPKPLNFPTNSKPSLDPSSTDTYSEQ	60	
DB	1	MNSPSPKLPIDKSHSLQLOPOSSASIFNSPKPLNFPTNSKPSLDPSSTDTYSEQ	60	
QY	61	DOEKGEKKKDTAFQTSFDRNFDLNSIDIQOTIQHQOQPOQOQOQLSQTNNLIDFSF	120	
DB	61	DOEKGEKKKDTAFQTSFDRNFDLNSIDIQOTIQHQOQPOQOQOQLSQTNNLIDFSF	120	
QY	121	QPTMTSLDTLKONPVDKVENHAPTYINTSPKSKIMKATPKASPKKVAFTVNTPEIH	180	
DB	121	QPTMTSLDTLKONPVDKVENHAPTYINTSPKSKIMKATPKASPKKVAFTVNTPEIH	180	
QY	181	HYDNRKEEDSOQKEDSVPLIOHQMDPSOFNSDSDTNAASVPPPLTKRTPFA	240	
DB	181	HYDNRKEEDSOQKEDSVPLIOHQMDPSOFNSDSDTNAASVPPPLTKRTPFA	240	
QY	241	QLLNKNEVNSEPALTDMLKR 263		
DB	241	QLLNKNEVNSEPALTDMLKR 263		
RESULT 2				
B90577				
lipoprotein b (imported) - Mycoplasma pulmonis (strain UAB CTIP)				
C:Species: Mycoplasma pulmonis				
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001				
C:Accession: B90577				
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer,				
Nucleic Acids Res. 29, 2145-2153, 2001				
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p				

QY 58 SEODQEKKEKKDPAFOTSPDRNFIDNSIDIOQTIOHQOQOPOQOQOOLSDTDNNLIDE 117
 Db 151 ILEEKANNTYNNDDYTKRPTTSEIQTKRPTTPOESTNIENSQOQPTP-----SKVDNVTVA 206
 QY 118 FSPOTPMSTLDTLTKONPTVDKVN---NHAPTYINTSPNKSIMKATPKASPKKVAFTV 174
 Db 207 TNPKEPVNVSKELKNP--EKLKELVKNNDTDRSTKPVATAPTSVAPKRLNAKMFAY 264
 QY 175 TNPETHNPDRNVEE-----EDQSOKEDSVEPPLIQHQKDPDSQFNYSDEDTNASVPTTP 230
 Db 265 AQPAA--VASNVDLITVTKOTIVGDKDVAAAHQDK-----IEYDEFTL----- 312
 QY 231 PLHTTKPTFAQLKNKNNEVSEPALTD 258
 Db 313 ---DNKVKKGGDTMTINTYCKNVIPSDLT 337
 RESULT 6
 S46660
 C:Species: Penicillium chrysogenum
 C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 26-Aug-1999
 C:Accession: S46660
 R:Prade, R.A.; Timberlake, W.E.
 Mol. Gen. Genet. 244, 539-547, 1994
 A:Title: The Penicillium chrysogenum and Aspergillus nidulans weta developmental regulat
 A:Reference number: S46660; MUID:94359480; PMID:8078481
 A:Accession: S46660
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-499 <PRA>
 A:Cross-references: EMBL:X80058; NID:g510873; PIDN:CA56364.1; PID:g510874
 C:Superfamily: regulatory protein weta
 Query Match 9.5%; Score 132; DB 2; Length 499;
 Best Local Similarity 24.4%; Pred. No. 0.22;
 Matches 60; Conservative 40; Mismatches 92; Indels 54; Gaps 14;
 QY 50 NSSSTFYSEDOQEKKEKKDTAFOTSPDRNFIDNSI-----DIOQT-----QHQQOOP 101
 Db 8 HSFNLFMOYVMETSAVDGKDSAL-SDFDLFPLDSLSDCGDLPPTVSTRKRGSPQ 66
 QY 102 QOQOOLSDTDNNLIDFSFOTPMSTLDTLTKONPTVDKVNENH---APTY---INTSP 153
 Db 67 WSEMSLDDGAAADHFAHFD-----TVHPSAISVDNLNLFVPSRPTASHGLSTSP 118
 QY 134 NKSIMKATPKASPKKVAFTVNPETHNPDRNVEEEDQSOKEDSVEPPLIQ----- 206
 Db 119 S---TPPATPRKRQSAI-ITPKSIRH---RSPMERKSHLRKQSFSPSLMRSSNLKSA 170
 QY 207 -----HOKKDPDSQFNYSDEDTNASVPTTP-----LHTTKPTFAQLKNKNNEVSEPAL 255
 Db 171 RMAVEAQAQRLQ-NFSLHGSSEDRPLSPSPSDVLIQHENMPT-EQIMNHGD-SAERAS 227
 QY 256 LTDMKL 261
 Db 228 QYDARL 233
 RESULT 7
 S57150
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Sep-1995 #sequence_revision 01-Mar-1996 #text_change 06-Feb-1998
 R:Rose, M.; Koetter, P.; Entian, K.D.
 submitted to the Protein Sequence Database, September 1995
 A:Reference number: S57132
 A:Accession: S57150
 A:Molecule type: DNA
 A:Residues: 1-1380 <ROS>

A:Cross-references: EMBL:Z49627; NID:g1015856; PID:g1015857; MIPS:YJR127c
 R:Thomas, D.; Barbey, R.; Surdin-Kerjan, Y.
 submitted to the EMBL Data Library, December 1993
 A:Reference number: S43751
 A:Accession: S43751
 A:Molecule type: DNA
 A:Residues: 'MHN', '18-1116', 'IF', '1119-1130', 'H', '1132-1142', 'S' <THO>
 A:Cross-references: EMBL:L26506; NID:g432497; PID:g432498
 C:Genetics:
 A:Gene: SGD:ZMS1
 A:Cross-references: SGD:S0003888; MIPS:YJR127c
 A:Map position: 10R
 Query Match 9.4%; Score 130.5; DB 2; Length 1380;
 Best Local Similarity 18.9%; Pred. No. 0.92;
 Matches 66; Conservative 53; Mismatches 121; Indels 109; Gaps 13;
 QY 3 STPSKLLPIDKSHSLDLOQSSASIFNSPTKPL-NEPRTNSKPSLDPNSSDPTVSEOD 61
 Db 331 NFPSSMHTKRRAHASFASSAMTYSSNSPHHSITNPELVEDAPH-QVGFSTPQMTAKOL 389
 QY 62 QEKKE-----EKKDTAFOTSPDRN-FDLNDSIDIOQTIOHQOQOPOQOQOOLSDTDN 112
 Db 390 MESVSELDLPPPLTLDDEPPQAIK--FNLNLFNNDPS-----GQOQOQOQOQOQNSTSS 438
 QY 113 NLIDFSFOTPMSTLDTLTKONPTVDK-----VNENHAP----- 146
 Db 439 TVVSNNGSTVATPPGVYLLSSGSLDYLTMNSAHAGAGCYSSHHSPDLGCFPSHDKPT 498
 QY 147 -----TYINTSP-----NKSIMKATPKASPKKVAFTVN 176
 Db 499 VSEFNLPSSEFNTIPENSTTASNSYSNLANOYTRQMSNEDPLMSLSPKNPPTVSDSSST 558
 QY 177 PELHNPDRNVEEEDQSOKEDSVEPPLIQHQ-----KDPQ-----FN 216
 Db 559 INFNPCTNNLLEPSMEPNKDSNIDPAIDDKLSEFINNSDPKSTFKINFHNDIGFT 618
 QY 217 YSEDTNASVP--PTPLHTT-----KPPFAQLKNKNNEVSEPE 254
 Db 619 YSPSSRSSTIPKSPPNHSTSLNHEKASLSPLNLNLNGSIDLPSTPQ 667
 RESULT 8
 S58147
 C:Species: Schizosaccharomyces pombe
 C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999
 C:Accession: S58147; T38551
 R:Gentles, S.; Churcher, C.M.
 submitted to the EMBL Data Library, July 1995
 A:Reference number: S58145
 A:Accession: S58147
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1087 <GEN>
 A:Cross-references: EMBL:Z50142; NID:g1052783; PIDN:CA90490.1; PID:g1052786
 R:Gentles, S.; Churcher, C.M.; Bartell, B.G.; Rajandream, M.A.; Walsh, S.V.
 submitted to the EMBL Data Library, July 1995
 A:Reference number: Z21799
 A:Accession: T38551
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1087 <GE2>
 A:Cross-references: EMBL:Z50142; PIDN:CA90490.1; GSPDB:GN00066; SPDB:SPAC27.03c
 A:Experimental source: strain 972h-; cosmid c2f7
 C:Genetics:
 A:Gene: SPDB:SPAC27.03c
 A:Map position: 1
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
 C:Keywords: ATP
 F:697-959/Domain: protein kinase homology <KIN>
 F:705-713/Region: protein kinase ATP-binding motif

Query Match 9.0%; Score 125; DB 2; Length 646;
Best Local Similarity 24.2%; Pred. No. 0.84;
Matches 54; Conservative 32; Mismatches 93; Indels 44; Gaps 9;

```
QY 48 DPNSSDTYTSSEQDEKKEKKDTAFQTSFDRNFDLNSIDIQOTIOHQOQPPQOQQL 107
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 432 DARNHLDVVKAAQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 481
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 108 SQTNNLIDERSFQTPMTSTLDLTKONPTVDKVENNHAPTYI---NTSPKKSIMKKATP 163
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 482 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 539
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 164 KASPKVAFTVNPENHYPDNRVEEDSQKEDSVPEPLIQHWKD---PSQFNYS 218
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 540 KANKK-----PQEQNNAVDDAQ---AGEPSAHREQSKDSQNGQPSGTGEQN 581
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 219 DEDT---NASVPTPTPLHTTKPTFAQLLNKNEVNS--EPEAL 256
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
      582 DEOSDKANAAPSTSVTTSDDPNLDPMLRLKLEOVESARDPAL 624
```

RESULT 15

```
F90537
liporotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: F90537
R:Chamblaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A9512; MUID:2126165; PMID:11353084
A:Accession: F90537
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-773 <RUR>
A:Cross-references: GB:AL445566; PID:q14089619; PIDN:CAC13379.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPU_2060
A:Genetic code: SGC3
```

Query Match 9.0%; Score 124.5; DB 2; Length 773;
Best Local Similarity 23.5%; Pred. No. 1.1;
Matches 61; Conservative 37; Mismatches 121; Indels 41; Gaps 11;

```
QY 10 PIDHSHLQLOPQSSASIFNSPTKPLNFPRNSKPSLDPNSSSDTYTSE-----QD 61
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
      37 PSKQNTKQTSPOSAPKE-----NNTNTNNSIISPQNPSSSKTPTETQVPTKPED 87
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 62 QEKGEKKDTAFQTSFDRNFDLNSIDIQOTIQHQ--QOPOQOQOQLSOTDNNLIDERS 119
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 88 QNKRPQSPKDPKIDKQCN-EGSKAPEIKMSQKQAPQVQKQPEDPKPKET----- 140
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 120 FQTPMTSTLDLTKONPTVDKVENNHAPY---YINTSPKKSIMKKATPKASPKVAFTVT 175
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 141 -QKPPVASEQNK-E-PQDPKAPKQAEPTPKDPQVQKDMANKNIQ---GPKV-PEKQAOQTPK 194
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 176 NPEIHHYPDNVEEEDSQKEDSVPEPLIQHWKD--SQFNYSDEDTNASVPTPTPLHTT 235
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 195 DPEIKNM-DQKQAPQEQKQPEASKKPETQKPTPNPSSNTQOENKQPEYOKTP---E 249
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 236 KPTFAQLLNKNEVNSEPEA 255
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 250 KPKAPKAEEPPQNGTQNNNA 269
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

Search completed: March 17, 2003, 12:27:09
Job time : 19.1863 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:19:45 ; Search time 58.7193 Seconds
(without alignments)
5839.012 Million cell updates/sec

Title: US-09-964-858-1

Perfect score: 8631

Sequence: 1 MNSPTSKLPLDKHSHLQLQ.....WVNLMLQQQQQQQQSSSQ 1664

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

otal number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_bacteriap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	382.5	4.4	3257	5 Q9V736	Q9V736 drosophila
2	381	4.4	2081	10 Q9LH98	Q9LH98 arabidopsis
3	369.5	4.3	2647	5 Q9U4X0	Q9U4X0 plasmodium
4	369	4.3	1723	5 Q8T224	Q8T224 dictyostell
5	367	4.3	704	3 Q9P7Y8	Q9P7Y8 schizosacch
6	355.5	4.1	1514	5 Q8SY55	Q8SY55 drosophila
7	355.5	4.1	2910	10 Q9FND5	Q9FND5 arabidopsis
8	346.5	4.0	2771	5 Q26216	Q26216 plasmodium
9	338.5	3.9	5327	5 Q76891	Q76891 drosophila
10	337.5	3.9	1738	5 Q76329	Q76329 dictyostell
11	337.5	3.9	16215	5 Q9NFS3	Q9NFS3 drosophila
12	336	3.9	6815	5 Q9I7U4	Q9I7U4 drosophila
13	335	3.9	2678	5 Q9NDS4	Q9NDS4 dictyostell
14	333.5	3.9	3111	5 Q9VH10	Q9VH10 drosophila
15	331.5	3.8	3147	5 Q17464	Q17464 caenorhabdi
16	331	3.8	2472	5 Q8T2M5	Q8T2M5 dictyostell

17	329	3.8	6713	16 Q99U54	Q99U54 staphylococ
18	329	3.8	6713	16 Q931R6	Q931R6 staphylococ
19	327.5	3.8	1807	5 Q8MSL1	Q8MSL1 plasmodium
20	325.5	3.8	2768	5 Q9VC00	Q9VC00 drosophila
21	325	3.8	1650	5 Q77328	Q77328 plasmodium
22	324	3.8	1302	2 Q49547	Q49547 mycoplasma
23	322.5	3.7	3354	5 Q8T101	Q8T101 bombyx mori
24	320	3.7	5085	11 Q9JKS6	Q9JKS6 rattus nori
25	320	3.7	17352	5 Q95YM2	Q95YM2 procamburus
26	318	3.7	1236	5 Q9GTX2	Q9GTX2 plasmodium
27	317.5	3.7	3130	5 Q9BK46	Q9BK46 plasmodium
28	316.5	3.7	1593	5 Q8T1T7	Q8T1T7 dictyostell
29	316.5	3.7	3254	5 Q9BK45	Q9BK45 plasmodium
30	315.5	3.7	1616	4 Q96PH3	Q96PH3 homo sapien
31	314.5	3.6	2867	5 Q9NPM3	Q9NPM3 plasmodium
32	314.5	3.6	3724	5 Q77320	Q77320 plasmodium
33	314	3.6	4498	5 Q9W223	Q9W223 drosophila
34	313.5	3.6	5147	4 Q9Y6V0	Q9Y6V0 homo sapien
35	312.5	3.6	1769	5 Q16625	Q16625 caenorhabdi
36	312	3.6	2038	5 Q967Y0	Q967Y0 dictyostell
37	311.5	3.6	1803	5 Q9VU16	Q9VU16 drosophila
38	310	3.6	1492	5 Q8SSU1	Q8SSU1 dictyostell
39	309.5	3.6	1271	5 Q25860	Q25860 plasmodium
40	308.5	3.6	5147	4 Q9Y6V0	Q9Y6V0 homo sapien
41	308	3.6	2454	5 Q8T2G3	Q8T2G3 dictyostell
42	306.5	3.6	2062	4 Q9H231	Q9H231 homo sapien
43	306	3.5	2151	5 Q9NG79	Q9NG79 trichomonas
44	305.5	3.5	1204	5 Q8T134	Q8T134 dictyostell
45	305.5	3.5	2752	5 Q9BAY0	Q9BAY0 plasmodium

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	3257 AA.
Q9V736	Q9V736		
AC	Q9V736		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	CG12864 protein.		
GN	CG12864.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,		
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,		
RA	Abdill J.F., Agdayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,		
RA	Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunth B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,		
RA	Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,		
RA	Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,		
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,		
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,		

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Mishina N.V., Modarity C., Morris J., Mostreli A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shie B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL: AE003814; AAF58230.1;
DR FlyBase: FBgn003946; CG12864.
DR InterPro: IPR000637; AT_hoek.
DR PRINTS: PR01608; BACINVASINC.
DR SMART: SM00384; AT_hoek.1.
SO SEQUENCE 3257 AA; 355970 MW; 7E9212C20A40810C CRC64;
Query Match 4.4%; Score 382.5; DB 5; Length 3257;
Best Local Similarity 19.6%; Pred. No. 3.4e-08;
Matches 373; Conservative 292; Mismatches 679; Indels 559; Gaps 88;
QY 4 TPSSKLLPDI-----KSHILOLOPOSSASIFNSPTKP-----INFPTNSKPSIDPNS 51
DB TGSSTNLPODDIEMASNHQETDLKCAPDRVALDKSESTPRVEEQGLKVPDPTDLDESK 647
QY 52 SSDT---YTSDDQDKGKEE-KKDAFOTSPFRNFDLNSIDIGQITQHQQQPPQQQQ- 106
DB 648 VSESAKNHIELEDDKKDKKEEQKESPNKSKETN---ENSVITVNEVELPAKKAKEA 704
QY 107 ---ISQTNNDIERSF-----OTPMSTLDLTQONPVNDVNNHAPT 147
DB 705 GNIVESSQALAEDEKLAELIAEYGVKEANEVSVTSVQG--EONPIYELVKELEDET 762
QY 148 YIINSP-----NKSIMKAPPKASPKVAFTVTNPEIHHPYDNRYEEDSOQKED 198
DB 763 ISEVPAONDOSSVEDQTLADKENVEKPSVKAPASS-----KDEPPAEENLPAPQD 816
QY 199 SVE---PPLIOHOMKD-----PSQFNYSDD-THASVPTPTPLHTTPAQQLLNKNE 248
DB 817 PIEQKTPVAKNQQHDKHNEAPAESLSVSDIPSSVTPSKRNHSSP--ANTPKKSKE 874
QY 249 V-----NSEPALTDMKLKRENFNLSL--DEKV-----NLYLSPTNNN 286
DB 875 IEALQSSVPRALASDKATPQNLRESRSKRLTKLTLLMDTMRSSPRIGRPAESHS 934
QY 287 SKNVSDMS-----HLOLQDASKNK-----TNE 310
DB 935 SHERSPEMKVTVSLAKDLTITIDKEKEIELKSLPDASETKDVKITTTTASDTSILDE 994
QY 311 NIHLSTALKAPKNDIENPLNSLTNADISLRSSGSSOSLSLRDNANVLESVQC--SPK 368
DB 995 NPSSSKTEMKKLCG--KPLK-----AKKMSRTSETEVKAIALDSN--EDIPSISFSIK 1042
QY 369 KVNGLSLNDIGKFSDEVEVSLPR--DLSDRLKLETTKHDAPEHNENFNDAKSNIT 425
DB 1043 CVEEHLTSSSEEQ--KDEKELLCPKPOIDCTNIDLE---OSTAIEDTDEO--VEEKRSNR 1096
QY 426 NKGOLVSS-----DHLDSFDSRYNHTEOSIINLNASOSQISLNALEKOR--- 473
DB 1097 RKSRRINERKFTETDTLSDHLDAKKAENSLAISMRKCTLETQOSPPYAKKKRNSGR 1156
QY 474 -QOQEOQOTQAEPEETS---PSDNKIKVQEPKSNLEFAVVTIKKEPVSAFETKAPRE 529
DB 1157 LSRREKSVINAASEKSPSAISOSTERQQLNEN-----PSKKD--KRTEOSGNKKE 1208
QY 530 FSSRIIRIKNEDEIAPADIIHPKKENEA-NSHVEDTALLKALINDEESDTTGNSTKMS 588

DB 1209 ---AVGQPLDKETSSSTNIIDKKSNESFDSAQPSD-----RLNOKESAFTKLSISSP 1260
QY 589 IREHISDMKL-----EDSNGDREDNDIRFEKSDI-----LNDV 625
DB 1261 KTKMKQDDKDLALSKGGSNPIRTPTGEDSRQTDKHQDNQDKHEEDSSKLKANDET 1320
QY 626 SQTSDIIGKYGNSSEIYTK--TLAPRSPDNDKENSLSLED-----PANNESL 673
DB 1321 KSSSEKDAEPIKSDSQDAKAPRLSKPSKNNKKKKKKKKPNDSIAESDIEGQVNTVEY 1380
QY 674 QOOLEP--HTKEDDSLANSNAP-----PEELTTPVEANDY 711
DB 1381 QATCSPSESNKKDMKVSQETNEEPMISETELGRIKKRQAFHIENPKD--DLHITPONEN 1439
QY 712 SSFNDYTKTFDAYSSEESISREHETDSKPINISIMHKOEKKKQIKQIKVPTKQIIAS 771
DB 1440 QSIAGY--NFEKQVPLPEST---ESDT-----PIKIKPTKTYLMCT 1475
QY 772 QOYKNDQESRVTSDKVKIPNAIQFKKKEVNVASRRVSPMDDLNVQFLPELSEDSGF 831
DB 1476 KNTSLSLASEDPDILPE-----QKLIITSKQDSNPDLDMAN-----NLETSS 1520
QY 832 KDLNFANYSNNTNRRPSFTPLSTKNVLSINDDPNVVEPPEP--KSYAETRNARLSANKA 890
DB 1521 QDPKEHEFSQDT-----FT--DNSDIIPSCTKKSQIVFTPTPKSSDQTKN-----SFI 1567
QY 891 APNQAPLP-----PORQPSSTRSNKRRVRRFVPE-----IRRS 929
DB 1568 TPNRSKSKRNYSKEAKRLDNSEESQNAASSSKVKOKELTPTYASCRLKRLVLRKP 1627
QY 930 SALAPCDMYNIDFDDGAGSKPTIK--AEGMKTLPSMD-----KQDVKRI 972
DB 1628 TSSLPTNSRKSIFKKTTPAKSKRLTILEMEKTPSEPSVSLGEVNPDPMAAREVAVL 1687
QY 973 LNAKKGYODEYINAKINDQPKKNSIYD-----PEDRYE-ELQO--TASINATIDSSI 1025
DB 1688 HESDDLESNEIPENEVEDETEASAEDDNKLRKEDDHELVNDICAAKSNPTDOST 1747
QY 1026 YGRPDSIST--DMLPYLSDELKRPITALISAD-----RLMEQEVNHLRS----- 1068
DB 1748 KQASSNKSPTSDSVLQETDELDS--NSLINATQGEDTPTKELTEEVPRNKKIVEDESKQ 1804
QY 1069 -----NSVLYHPGAGAAATNS-----SMLEPP-----DEFLINSPAR 1099
DB 1805 ELIKDLEPDNALIEDTASTAKAAEEMDLYIKESNVASVLAEPDYTDDBELQSPRP 1864
QY 1100 NVSNNSDVVAISGNASTISF-----NOLDNPNPDQAT-----IGQYIOEOP 1140
DB 1865 N-SETTSTVTDDEPSTSSVYKRSILKRRPADSSQDDEAKRRQODVEKSLTGKKEQYKP 1923
QY 1141 A-----SKSANTVNG-----DDGLASAPETPTPKKESISK 1174
DB 1924 ARRQLAVEERPSLKRKTESEAKSYQGYKISTIGNETIMSTTAIRIETNREAASTS 1983
QY 1175 PAKLSA-----SPKSPRIKGSFVYRIKKGSIAGIEPIPKAT--HK 1215
DB 1984 PSARKSAVOEAKHVEYTHIILGPPGKRLHSDSPAALVKK-----PMYOTLLSTLSLOK 2039
QY 1216 PKKSQGNIEISNHKVR-----DGGISPSGSGEHOONHNSMVSVSQYTDATST 1263
DB 2040 PSTLDDGSP--KIRKSLKSIADENIDGQOSIFSSSSVLNKNYSVVA--PRKVNISVSL 2095
QY 1264 VPDEKNQVQ-----HKPREKOKONNNNNNNNNNNKOKTDIRGVVDDKIP 1307
DB 2096 L--QSKDTQVETLAASSSTETPLTKKEKLTQKTSKPKRGNNKTESKKSLSVQG-----P 2147
QY 1308 DVLQERKGLFERYLGIKNIN-LPDINTNKGRTLLTDNGYVCVUTPREYNNMDHNVAIGK 1366
DB 2148 QMKTQKSEB--AVSGPILKYLKSETESSRKIVSYVTG-----RKQIQG 2190
QY 1367 EFELTVADSLFILTAKASYKPRGTLVEYTEKKVNVASRNLSRFGSKDITTTTKVEYPT 1426

Db 2191 -----LEVL--KKPESKSESLVEAISRKKSOYORLSKIDGRKSEGTSLPQPDV 2239
OY 1427 EVKDTMANKFAPDGSFARCYIDQOEEDQITGASQFDLNCFNEMETMSNGNPMKRGK 1486
Db 2240 SKSETALKALP-----KETEFPVO-DAETKMSKGRGHQNAVK 2278
OY 1487 YKIAQLEVMKLYVRSDPREILPTSIKRSAY-ESINELNNEONN 1528
Db 2279 TKTEQ-----PKSKPK-----TEVRSLOAEATLMDMS 2309

RESULT 2

O9LH98 PRELIMINARY: PRT: 2081 AA.

AC O9LH98: 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
eurosids II: Brassicales: Brassicaceae: Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLIMBIA:
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.,
RL Submitted (Mar-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLIMBIA:
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.,
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety p1,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL: AP002057; BAB03174.1;
SO SEQUENCE 2081 AA; 232851 MW; D3603E1F85EFFE29 CRC64;

Query Match 4.4%; Score 381; DB 10; Length 2081;
Best local similarity 19.0%; Pred. No. 2.3e-08;
Matches 295; Conservative 264; Mismatches 627; Indels 364; Gaps 61;

OY 47 LDNSSDVTYSQDQKGEKKKQTA-----FQTSFDRNFDL-----DNS 87
Db 460 LKGNASVEAKT--NNSSSEKREESQSRSENYMKNKETTGGENVNIGESIGDSTKONS 516
OY 88 IDIQOTI-----QHQQOOPQOQOOLSDTDNMLIDFESFQTPMTSTLDTKONPTVDKY 140
Db 517 LEKKEVKKRVKVDANESDGNSTKRRHQEAQYNNGV-----STEDKNLDNIGADEQ 565
OY 141 NENHAPTYNTSPNKSIMKATPKASPKVATVTPETLHHYDNRVEEDQ-----QOK 196
Db 566 KKNDKSVETVNTDGDHTKEKR--EETOGNGESVKNENLENKEDKELKDESGAKTN 623
OY 197 EDSVEPPLQHQMKDPSQFYSDPT--NASVPTTPPLHTTPTPAQLINKNEVNSP 253
Db 624 ETSLEKRBQTOGHDNSINSKIVDNKGGNADNKEKEVHVGST-----NNNNMSKE 677
OY 254 EALTDKLK-----RENFSNLSDKEVNLVLTPTNNNSKNVSDMSHLQ--- 298
Db 678 DTKSEVEVKKNDGSSKEGEGKKNKDSMDKLENKESQTSKDKSVDKOBEAOIYG 737
OY 299 -----NLQDASKNTNEN--IHNLSFALPKAKNDIE----- 327
Db 738 GESKODKSVYAKKKESKKNKKTJTNENVRKKEENVQGNKKESEKVEGKESKDAK 797
OY 328 -----NPLNSLTNADISLRSSGSSOSSLSQSLRNDNRVLESVPGSKKKNPGLSLNDGT 380
Db 798 SVETKDNKKLSTENDEAKERSGEDN--KEDKEESKQSVYAEKKNENGVDTNVGN 854

OY 381 KGEFDEVESLPRDLSRDK---LETTKEHDAPENNE-----NFTDAKSTN 424
Db 855 KEDS-----KDLKDDPSVEYKANKESMCKKREYVRNDKSSKTEYVRDANNDID 905
OY 425 TNRGOLLVSSDDHLDSEFRSYNHTEOSILNLSASOSQISLNALEKOROTOREOTOAA 484
Db 906 VQGG-----SGESVYKKYKDEKKEGKNEKNKDTINTSSQK-----GDKKAKKESKNSNM 957
OY 485 EPEE-----ETSPSDNIKVAQEPKSNLEFVAVTITKKEPVSAETELKAP 526
Db 958 KKEDEKKEYVNNELKQEDKNETTKSENKLEKNENKNE-----KKESEDSASKNRE 1011
OY 527 KREFFSRILRIKNE-----DEIAEPADIDHPKKEENANSHVEDTDLK--KALNDE 576
Db 1012 KKEYEKKSKTKKEAKKKEKKKQKKEEDSEKAKKKEKESKRDIAKAKKEETKEK 1071
OY 577 ESDTTONSTKMSIRPHIDSKLEDSNDGREDNDISREKSDILNDVSQTSIDIDKY 636
Db 1072 ESENHKSCKKEDKKEHEDNKSMMKEEDKKEKKHEESKSRKKEEDKDMEDLED----- 1125
OY 637 GNSSETITTTTLAPRSDNNDKENS---KSLDPPANNESLQOOLEVYHTREDSILANSS 693
Db 1126 QNSNKK-----KEDKNEKKKSOHVKLKESDKKEKKEKSEKSETEIEISSKQKN 1176
OY 694 NIAPPEELTPVVEANDYSSFNDVTYKTFDAYSSFEESLSREHETSPINFTSIWKQ-- 751
Db 1177 EVOKKEK-----KSKDQOKKKEKEMKSEBEKKLKNEDRKKQTSVEENKKQKE 1226
OY 752 ---EKOKKHQIHKVPTKQ-----ITASYOQYKKEQESRYT--SDKVKTIPNALQFKFK 799
Db 1227 TKKEKNKPKDKKNTTKQSGKKEKSEKSEKAEENOQKSOATTQADDESKNEILMQADS 1286
OY 800 EVVVMKRRVYSPMDLNVQFLPELSEDSGFDLANANSNNTNRRSTPLSTKTVLS. 859
Db 1287 QADSHSOSQAD-----SDESKNEILMQADSOATTQRRNNEEDRKKQTSVAE 1331
OY 860 NINDPNNVVPPEPKSYAETIRNARLSA-----NKAAPNOAPLPQPOPSSTRSN 910
Db 1332 NKQKQETKEKKNPKD--DKNNTTKQSGKKEKSEKSEKAEENO-----QKSATIQAD 1383
OY 911 SNKRVSRFRVPTFEIRTSALAPCDMYNDIFDFGAGSKPT-----IKAEGMKTLPMSD 965
Db 1384 SDE-----SKNEILMQADSOADSHSDSQADSESKNEILMQADSOATTQRNN 1430
OY 966 KDOVKRILNAKKQTOQEYINAKLVQDKPKKNSI-----VTPDEDRKEELOQTSH 1017
Db 1431 EEDKKQTSVAENKKQKETKEKKNPKDKKNTTEQSGGKKEKSEKSEKAEENOQS--- 1487
OY 1018 NATIDSSYGRPDSISTDMLPYLSDCLKPPTALLSADRLEFMEQEVHPLRSNSVLYVPGA 1077
Db 1488 QATTQGE-----SDESKN--EILMQAD-----SQADTH--ANS-----QGD 1519
OY 1078 GAATNSMLPEPDEFELINSPARVSN-----SDNVAISGNASTISTFNOIDMNEFDQAT 1132
Db 1520 SDSESKNEILMQADSOADSOQTDSDSESKNEILMQADSOQTDSDSESKNEILMQADQAKI 1579
OY 1133 GQKIQEOPAKSANTVAG--DDDGLASAPETPRT--PTKKESTSSKRALTSASPKKSPIT 1190
Db 1580 GESLEB-----NKVAGKEDNGDEGVKENSKTLEVGRHEESKQGTNENGGEVSTEE 1632
OY 1191 GS--PVYRIKKNKSIAGIEPIPKATHPKKSPQ--GNEISNKK--VRGGGSISSGSSHQ 1245
Db 1633 GSKDSNLYERNKNG--GKEDSTKESSEGGKTYEINGGELSTEEBSKQKI--EGKSGKE 1667
OY 1246 HNPMSVSPSQYTDATSTVDENKDYONKPREKOKKNNHNNHNNHNNKQTDIIPGVVDE 1305
Db 1688 NSTKESGKDKIEGK-----EGKE-----NSTKESGDKKINETIGDKEATMEGSKDGG 1738
OY 1306 IPDVGLOERKGLFRVVGIGININILPDIINTKGRFTLLDNGVHCVTTTPVPMNDHNAYI- 1364
Db 1739 TNSTGDKSKSKSVYINGVAKDSDLKD--DSKNGDIN--EINNG-----KESDVKNQVTEIQ 1790
OY 1365 GKFEFLTVADSLFELITLKASYEKPRGTIVEVTEKKVYKSRNRLSLRIFGS 1414

```

Db 1791 GNDNLSLTJSTSE-----PBGDKLD-TNKDSMKN-WTMEAGGS 1827

RESULT 3
ID 0904X0 PRELIMINARY: PRT: 2647 AA.
AC 0904X0:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Putative erythrocyte binding protein EBL-1 (Fragment).
GN EBL-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=20078864; PubMed=10613703;
"Peterson D.S., Wellem's T.E.;
"EBL-1, a putative erythrocyte binding protein of Plasmodium
falciparum, maps within a favored linkage group in two genetic
crosses."
RT Mol. Biochem. Parasitol. 105:105-113(2000).
RL EMBL, AF131999; AAD3018.1; .
DR NON-TER 2647 2647
SQ SEQUENCE 2647 AA; 304550 MW; AE98F88FD754E300 CRC64;

Query Match 4.3%; Score 369.5; DB 5; Length 2647;
Best Local Similarity 19.5%; Pred. No. 9.5e-08;
Matches 383; Conservative 305; Mismatches 717; Indels 555; Gaps 95;

OY 21 POSSSASIFNSPTKPLNFPRTNSKDS-----LDPNSS-----SDTYSEDOQ--- 62
Db 825 PSSSKAL---KPKIKDVFIEETKSELSLTDKSKNTPNSSGGNGYDQRIKRDVAH 881
OY 63 -----EKKEE---KKDTAFOTSFDRNFDLNSIDG-----QTOHQOQ 99
Db 882 DCPKVEKSGEKEVPKIDAAVKT---ENFTSNRNDGEEKSGKGDHSSVPHSKDIKEEP 938
OY 100 QPOOOOQLSQTNNLIDERSFQTPMTSLDITKONFTVKNBNAPTYINTSPKSIKK 159
Db 939 QKVGSENLPKIEKM--ESSDSIPIT-----HIEAKGSSNSNDMD 978
OY 160 KATPKASPKKAAFTYINPEIHHYPRDKVEEDOSQOKEDSVPRPIQHQKDS--QFNYS 218
Db 979 PAVVSGRESKDVNLHTSERIKENEBCVITDDSKSIEIS-----KIPSDQNNIS 1028
OY 219 DEDTNAVSPRTPLHTTKPTFAQLLNKNEVNSEPEALDMLKRENSNLSLDE--RV 275
Db 1029 DLSQMAN-----EDSNQCKKETINPPS--TEKNLKEIHTKTSDDHGSKI 1072
OY 276 NLYLSPTNNNNSKNVSDMDSHLQNLQDASKNTNENIHLSPALKAPKNDIENPLN--SL 333
Db 1073 KSEIEPKELTESPLTDKKTESAIGD---KNHESVKSADIF---QSEIHNSDMRKI 1124
OY 334 TNADISLRSSGSSQSLQSLRDNK-----VLESVPSPKKNVNGLSLNDGKGFSE 386
Db 1125 VESVAVQDSSGSMST--ESIRTDNKFCTSEDIAPISNGHEK---IGSSADD--RGSED- 1177
OY 387 VESLPLRDLRDLKLETTHEHAPRENNENFIADKS--TNTKKGLVSSDHLD----- 439
Db 1178 --KSIIDKDSFENNNKSHSDIKQSDNCGSTDYESLTSPKGLDESVPSSIDMLKP 1235
OY 440 ---SFDRAVNHTEQSLNLNLSASQSLSLNALEKORQTOEOEQTOAPPEETSFSDNI 496
Db 1236 NKSPVTSFSDHVDSPNISLQASQN-----ADSYGGEKPSNLIKTDD 1280
OY 497 KKKQPKSLKLELVKTIKEPVSAIEIKAPKREFFSRILRIKNEDEIAPADIHPKKE 556
Db 1281 VSEKE-----KISVSPPNVSVTYDEGDKRQ-----GISDSSSIHHEIDPE 1320
OY 557 ANSHVEDTDLALKALNDEESDTQNSTKMSIRPHIDSKWL-----EDSND--GOREDN 610

```

```

Db 1321 KNLHVE--SFISGGL---EEGDIEKEKCK-----EDGSLIPISPEKINDFGKRENI 1367
OY 611 DDISREFE-----KSDILNDVS--QTSDIIDGKYG-----NSSEITTKLAPRSDNDKE 659
Db 1368 VDBSVSERVDNKRSDISNVASEQSSITISRRGTGETINNSELSEETSVDVNRRDH 1427
OY 660 NKSLEDPANNESLQOOLEVPTKEDDSLANSNIAPEELPLPVEANDYSSFNVDYTK 719
Db 1428 NMO--ENLVSSITQOSEEREKREKENADSSHELSLSIEVGETIRRDADAESEND--K 1483
OY 720 TPDAYSFEESLSREHETDSKP--INFISIMHOKOKKQIHKVPLKQIITASYOQYKNEQ 778
Db 1484 GBDITLOSEDOQIVTEKTKLIEPTVNL-----QPSPLIEIHKESSEET-----DDKDMT 1533
OY 779 ESRVTSDKVKIPNALQFKKEVYVWNSRRVVSPPDDDLNVSOFL--PELESDSGFKDLNPA 837
Db 1534 GGEVHTHTDA---NTROYRNSSEVEYITNKPDMEEMTTKPPSOYIEKEIETIDSTKN---- 1586
OY 838 NYSNNTNRPRESFTPL-----STKNVLSN-----ID-----NDPNV 867
Db 1587 --QDNDEQNSIIPLINKNNEEDGVSIPSTRVWMSGSVSRNEQIIEKDKHITDDPT 1644
OY 868 VERPEP--KSYAEIRARLRSANKAAPNPAPLPQROQ---PSTRSN--SKRVSRFRVP 921
Db 1645 INPSENGLKGYEMPN---DSIKSVTITESPLRDVEOMIEPIDGKGNKNNIIEPQES 1700
OY 922 TFEIR-----TSSALAPCDMY-----NDIF---DDFGG 948
Db 1701 TTEIRQOMGPSNVNIPPELIPVAGSKYLEAKERSMDADKGTITEDIYVEEDPNGIG 1760
OY 949 SKPTIK-----AEGMKLPDMSD--KDVYKRLMAKKGYTODEYINAKLVQDPKKNSTIV 1000
Db 1761 EHONLKEVHEQASELTYNSLDGRTVWEKERLDEPQGISIPNDRIITTEHIEDEKE--I 1818
OY 1001 TPEEDRYEEL-----QOTSINATIDSSIYGRPOSISDMLPYLSDLKKRPTALLSMD 1055
Db 1819 HEP---NELDAHNGEOEEMIRNEVSDNRM---DEQISDKNETROLQNDHEDSDODEITD 1871
OY 1056 RLFMEO--EVHPLRSNSVLHP-----GAGAATNSS-----MLPEPDP 1091
Db 1872 KREMENTLEENPNSSSDSLENPGEKKEGKTEHHSSELDVSVDYKYKGIENQITHEI 1931
OY 1092 ELINSPARVNSNNSDVAISGASTISFNQOLDMNFDDQATQOKIOEOPASANTVKG 1151
Db 1932 ESVEQODTNIPIGNSKETEDN-----SRADMEEKDVKIKIQTIE-----S 1973
OY 1152 DGLASAPETPTPTPKESISSKPAKLSSASPRKSPIKT-----GSPYVNI--K 1198
Db 1974 EEELEISKDTTSHSEKPSIEQSVNIYDS---KNEINQIEKNVONQONEDDPITLLEE 2030
OY 1199 KNGSIA-----GIEPIK 1211
Db 2031 OKNKIATILEQNKEXNPNPQSLSHKERTLLEVDLEQSMOTNCLTSELNKKCDSIQIPE 2090
OY 1212 ATNHKPKKSGFQGEISNHNKAKRDGIGSPSSGSEHQONPNVSPVSOYTDATSTVPEBNKY 1271
Db 2091 ASNTDKR--LNRDITENKQDFSEIEKSVGEIH--ENKDKLNNESAESD--DVPOQNK-I 2143
OY 1272 OKKPREKOKNNHNNHNNHNNKOTD---IPGVUDEIDPDVGLQERGLFFRVLGKININ 1328
Db 2144 EHDSEAGVITDOYETRTVEEKEKVFVGLPKPREERBDV---TRG-----S 2188
OY 1329 LP-DINTHKGRETTLLDNGVNCVTPPEVNMDDHNAIGKEFELTV-----ADSL 1376
Db 2189 VQGSYDAFEKLIHERLDSN-----EKKKEELANS--KEIERYVGLDITGIDSESBSDV 2239
OY 1377 EPLITLAKAYEKPRGTLVUTEKKYVKSNNRSLRFGSDIIT-----TTTKVPYTEK 1429
Db 2240 TI-----YKRPTDFEVNHH--VLSHD--IKSLFENKTVGLNEOKKETEKNIAGMI 2288
OY 1430 DTWANKFAPDGSFARCYIDL--QOFEDQITGKASQ--FDL-----N 1466

```


[illegible]

RESULT 5					
09P78	ID	09P78	PRELIMINARY;	PRT;	704 AA.
AC	09P78:				
DT	01-OCT-2000	(TREMBLrel.. 15, Created)			
DT	01-OCT-2000	(TREMBLrel.. 15, Last sequence update)			
DT	01-JUN-2001	(TREMBLrel.. 17, Last annotation update)			
DE	Hypothetical 78.7 kDa protein.				
GN	SFAPY07.03c.				
OS	Schizosaccharomyces pombe (fission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
CC	Schizosaccharomycetales; Schizosaccharomycetaceae;				
CC	Schizosaccharomyces.				
OX	NCBI_TaxID=4896;				
	[1]				
	SEQUENCE FROM N.A.				
	STRAIN=972H:-				
RA	McDougal R.C., Rajandream M.A., Barrell B.G., Saunders D., Harris D.;				
RL	Submitted (JAN-2000) to the EMBL/Genbank/DBD databases.				
DR	EMBL; ALJ36521; CAB66312.1; -				
DR	InterPro; IPR001849; PH.				
DR	Pfam; PF00169; PH; 1.				
DR	SMART; SM00233; PH; 1				
DR	PROSITE; PSS0003; PH_DOMAIN; 1.				
KW	Hypothetical protein.				
SQ	SEQUENCE 704 AA; 78716 MW; 755E28CD67F127AE CRC64;				
	Query Match	4.38;	Score 367;	DB 3;	Length 704;
	Best Local Similarly	21.0%;	Pred. No. 2.4e-08;		
	Matches 189; Conservative 140; Mismatches 315; Indels 256; Gaps				
Oy	767 VKPIPAIQFKKK-----EYNWMSRRVVSPPDDMDLVNAGQLPELSSEDSGEFKDLFNAYS	840			
Dd	21 LRIISPISPDYIECSDYASTIASITASIRESTRNRNNRNSNISSTAPEAFESDAEDGGSPFD	80			
Oy	841 NNTNPNRSF-----TPLSK-----NLNSINDNDPVNPVEPKPSVAETIRNARLSA	887			
Dd	81 OTLSNSSFFDHQSLLPFSEVRRTPTIYSMNATDSSSTSV-----DVNKENIIISL	132			
Oy	888 N-----KAAPNQADLPDPOROPSSSTRSNSKRARSREVPFTFEIRRTSSALACPDMYNDIF	942			

Dd	133	NDSCILIKSDEA-----					164
Qy	943	DDFAGSGPRTIKAGMKTLPSMDKDDVKRIIAKAKGVTOXEYINAKLVQDKPKKNSIYTD					10020
Dd	165	-NOGHGDIPI-----					177
Qy	1003	PEDRYEELQOTASIHNAITDSSITYGPRDSTIDMLPYLSDELK-KRPALLSADRLEPMEO					10611
Dd	178	-----PARVSCKSL-----					210
Qy	1062	EVHPELRNSVYLHVGACATNTSSMLPEDPFELINSBPARYVNSNSNVNVAISGNAIPSENO					11211
Dd	211	PTH-----					244
Qy	1122	LDMMFDDOATIGOKIOEQAPASKASANTVRODDGLASAPRTPTPT-KKEISISKEPAKLS					11800
Dd	245	-SFGFSEBSSSFODIKTRPRLSPADENR-----					293
Qy	1181	ASPKSPRIKIGSPRVYIKKKKSGIAGIEPRIPKATHKKKKSGFQGNELISNKKVNRGSGISPSG					12400
Dd	294	-----SITSIDSP-HUJEN--APRLPLPKVYSLP-----					320
Qy	1241	SEHOHNPSWVSPSOYTDTATST-VBENKDKVOYKREKOKNNHNNHNNHNNKOTDIP					12990
Dd	321	-----DPREFNVLSAPDALTPTTLRKNSKVVHNATSKOEMQTSRRVUNSCWMPES-LS					373
Qy	1300	GVWDEIIPDVGLOERCKLFFRVILGINIMLPRDNIHKGFETLTDGVCYCHTTPREYMNDD					13590
Dd	374	RNLSSSIQOQG--GSGKLFYRLMEIRNLTPRLASGMTTFEYTI--SGKH-IQVP-WNHLH					428
Qy	1360	HNVAIGKEFELTVADSLFELITLKASYEKERPGTGLVEVTEKKVAKNSRNLSPFLSGSKDIT					14190
Dd	429	STKIENEYFDESIISSIVCTLRAAYDPK-----					473
Qy	1420	TTKVEPPE-YKOTPMANKFAPDGSFARCYTDDLOQFEQITGKASQFDLCSFNMW-ETMSN					14760
Dd	474	KRKSMTTIDPVSEALHGFVSDEGTGEVYJINTDSVSFTALGRQOSMVLPIPMKKMTVDPAK					533
Qy	1477	GNOPMKRGKRPKIKOLEVVKMLVYVPRSDPREI--LPTISIRSAVESINELNNEONYFEGL					15340
Dd	534	DVKPL-----PRKVELEIHNFILP-ALPYSIKELPRAISISAMYDKLAEWORTLLCDSYL					588
Qy	1535	HOEGSDCEPIFKKREFKIMGTSLASHSEISHKTRAKINNVVDL---YVDKENIDRSN					15900
Dd	589	COQGGDCDPYMRBRRYFOLIGSKLVAFOOFSKVRATIDLSSEATHIYDDNHYSDOELE---					645
Qy	1591	HRNSVLDLNDHAKIKFANGCELLDPCARKNHMKIWIJNOEIIYRNFRQRPVNVNML					16500
Dd	646	-----GYLVEFSGRIIFSGMDIYIETAGTGBKDEMSYTLQNHIGDOCSWNNKNTKSGFL					700

RESULT 6		
08SY55		
ID	08SY55	PRELIMINARY; PRT; 1514 AA.
AC	08SY55;	
DT	01-JUN-2002 (TEMBLrel. 21, Created)	
DT	01-JUN-2002 (TEMBLrel. 21, Last sequence update)	
DT	01-JUN-2002 (TEMBLrel. 21, Last annotation update)	
DE	GH09355p.	
CN	CG6004.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxId=7227;	
RN	[1]	
RC	SEQUENCE FROM N.A.	
RC	STRAIN=BERKELEY.	
RA	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,	
RA	Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,	
RA	George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,	
RA	Miranda A., Mungall C.J., Nuno J., Pacleb J., Paragas V., Park S.,	

QY 124 MTSIDLUTK-----QNPVVDKVNENHAPTYINTSPKNSIMKKAIPKSPKVAFTVTPNPEI 179
DB 850 V-SLUVKSVQWOKPRL-----SPSEVSEETSKTVDKIEEK--PEE-----EV 892
QY 180 HHPRDNREEDDSOQOEVEPPLIOHQWKDSQFVNSQEDTNASVPPRLPHHTKPTF 239
DB 893 TLVQEOGVDSYGLFETKEEVSVP-----ESTELEPOQOEERSVJDPPLQKPLLESP 945
QY 240 AOLLNKN-----EVNSPEALTD-----KLRN 265
DB 946 SEVLEESSKTVDKIEEKTDSIELGEIAOEERSVTDLPLQEESQOPNEQETKLEKHE 1005
QY 266 FSN--LSLDEKVN-LYLSPTNNNSKNVSMDSHLQNLQDASKNKNTNENHNSPALKAP 322
DB 1006 PTNEEKSDVEVLVASPSKELEGETVVE---AENIENIKENEEOAKLOKSLSEV 1061
QY 323 KNDIENPLNSL-----TNADISLRSSGSSSSLOSILRNDN---RYLESVPGSPKKV 370
DB 1062 QT-VESPSILFSSEBODHTVAEIVDEKAKEEVPMLQIKNEDDATKIHTREVEQARDI 1120
QY 371 NPLG-----SLNDGKGFSEVVESSLPRDLSRDKLETKH-----DAPEH--- 412
DB 1121 GBSLFTICISIN---ONQPEOVKEACSEBQEKESTNSBNIVETALHSVAAEBETA 1177
QY 413 -NNENFLDKSTWT-----NKQOLVSSDDHLDSPRSTNHTBQSLINLLNSASOS 462
DB 1178 TNGESLDDVETKSVLLEVRKEEEMKTDAPRIDAIEKELELVKTVVQADAKIVNNE 1237
QY 463 QISLNALEKORQTOEOEOQOAEPEEET--SFSNDIKVKOEPKSNLE---FVKYTIKPEV 518
DB 1238 ETTAHESELSKJGDNHOKN--APVEATQNLDAEOLISRVVYVDTREADITKIEKVQE 1295
QY 519 SATETAPKRRERSSRLRIKNE-----ELAPADHKKKEANSHEVDTD----- 565
DB 1296 GPVITETP-----TIQGEIESETSLEKEEVD--OSSKDEHEHVLREDIPQCE 1345
QY 566 -----ALLKALNDEESDTONTKMSIRPHIDSMK----- 58
DB 1346 LKAAVOTSTVEBAATLKLTETNISPEAMHSETSLDLK--VDKEKAEVTVVIFSMNE 1403
QY 599 -----LEDSNDGREDNDISREFKSDILNDVQSOTDI----- 631
DB 1404 VGTSDAQAEFGFHTPCSESEIKDESQSEE---SVEVSKETVQESSEKDVNMILDVQ 1460
QY 632 --IGKYGNSSEITTKTAPPSDNDK-----ENSKLEDPANNESL-----QO 675
DB 1461 SGESEKYQENEDISLVS---KTENGDKFEELIPVVEGAGLDETHNQLLDVESVVKO 1516
QY 676 OLEVPTKEDDSILANSSMIAPPELTLPVNEANDYSSNDVTKTDAVSPEE----- 729
DB 1517 SLDTPEEETSKTIDKIEDKPEEVTLLH-QOEGREGSYGLDQTD-EAVSVLESRELQEO 1574
QY 730 -----SLREHETDSKPINFTISIMHOKOKKHQIHKVPTKOIILASYOQKNEQ----- 778
DB 1575 POOEELCLANQENETK-----LQEOVDKHE---PTKEEVSNDQSSPELELSNEY 1622
QY 779 -----BSRVTSKV---KIPNAIOFKKFKEVNWSRRVVSPPMDOLNVSQF 821
DB 1623 IOVSSASLSEGEPEYETVVEAEKIGEOVADKIQ-KSFETGEIVAEHASSLPSSEKEHET 1681
QY 822 LPELSEDSGFKLNFANYSNNNTNRPSPFTPLSTKNV-LSNINDPVPVPEPEKSAELR 880
DB 1682 VSEKIDDEKVAE-----PIGDMREGLDIAETTHLSLVSVOKEEVDIHIPVALPID 1737
QY 881 NARRLSANKAANOAPPLPQOPSPSTRSNKRVSRFVFPFELRTSSALAPCDMYND 940
DB 1738 EQEKTSTEGFETKSEAEADKRDDEHVOSTSPMLSEKNDNTQNSKTSIEDV--CMQOE 1795
QY 941 IPDDGAGSGPITKABGAKMTLPSMDKDVKRLNAKKVT-----ODEVINAKIV--- 990
DB 1796 -----SGTLEVVPKPEESKEDKSOEISETIEIEATSDQTLPIETSHNTDNTLSSELVSO 1849

QY 991 -DOKPKKNSIVTDEPDRYEELQOTASTIHNAITDSSISYGRPDISTDMLPYLSDELKRP-- 1047
DB 1850 DDQSPKKEVEIHHEEPKEADVEATSERNLVETS--DADNTLSQLVSETKREHKLQAG 1907
QY 1048 ---PT-----ALLS-----ADRLFMEOEVLNPLRSNVLVHPGA 1077
DB 1908 ELLPTEIIPRESSDEALVSLASREDDKVALQDNCADVDRETNDIOEERSISVETEESV 1967
QY 1078 G-----AATNSMLPEPDEFELINSPARVNSNSDVAISGASTISFN 1120
DB 1968 GETPKREHEDEIDAVHETPTARILLIENDSETLIAEK-----KNEELNETEKVALD 2022
QY 1121 QLDNFDQDA-----TIQKIOEQPASKSANTVRGDD---GLASAPETP-----PT 1164
DB 2023 HEEFVNHEAPKLEETKDEKSOEIPETAKATETITDQTLPIGTSQADQTPSILVSDKDDQT 2082
QY 1165 PTKESISSKPAKLSSAPKSPKPIKIGSPRVYIKKNGSIAGTPIKATHK----- 1215
DB 2083 PKOVEL-----LEETKETHKVOAEDIFST 2108
QY 1216 ---PKKSPQNEISN-HKYVDGSGISPSG-----SEHOQNPSPVSPSQYTDATSTVPD 1266
DB 2109 ETVPKESFTIAPVSMLASGDEPVTPOEGDYANTQOEERH---VSAET----- 2153
QY 1267 ENKDVOKPREKOKKHNNHHNNHHKOTDIPGVYDE-----IPDVGLQ----- 1312
DB 2154 EEKVGETKRESQAE-----GAEKSD--DQVEDESKTKTDVEVAGLENDVPTBEA 2201
QY 1313 ERGKLFPRVIGIKINIMLPDINTHKGRFTLLDNGVCHVTTPREVM-----DDH 1360
DB 2202 EHDEETYSILPVVGT-LTQLOT-----TLETERAINDSSASSEVSMKPEADQBEKKGDV 2255
QY 1361 NVAIGKEFELTVADSLFTLLTKASYEKPRGTLVETK----- 1399
DB 2256 VESNEKOF---VSD-----ILEAKRLHGDKSGEAEKIKESGLAGKSLPTEETINLOEENKE 2308
QY 1400 --KVVASRRNLSLRFSKSDIITTKFVPT-----VKDTWANKFADGSE-----ARCTI 1447
DB 2309 EVKVOETREIAQVLPREELIASSPLSABEOEHVISEDKEEREPQOOFNSTSEKISL 2368
QY 1448 DLOQFEDQITGKASQFDLQCFNEMETMSNGO-----PYMKRGKPYKIA----- 1490
DB 2369 QVNEHLKOFETSKKEQND---ETHETVKEEDQIVDIKDKKKDEDEEIVSSEYKKONKKA 2424
QY 1491 -OLEVKNLVPRSDPRELLPTISRSAYESINELNBNQNNYFEGYLHQEGGDCPIFKKRPF 1549
DB 2425 RELEVGNDRVSRDGEKEVPHNALENEEBMEVYASEK-----QISDPGVYIKKA-- 2474
QY 1550 KLMGTSLLAHSELSHKTAKINKSVYVDLIYVDKENIDR-----SHHRNFSOVLILDHAF 1604
DB 2475 -----SEAEHE-----DPAVDI-----KSNDDDFPTQOAPKDDSDSEVSADETV 2513
QY 1605 -----KIKFANGELIDFCAPRNH-----EMKIWQNLQOETIYRRNRPQO 1643
DB 2514 PREAIGEELKAVSSKVLDDIOENSNTAEVTFNADRLPQONLSSELQSHQSTQO 2567
RESULT 8
ID 026216 PRELIMINARY; PRT: 2771 AA.
AC 026216:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE Rhoptery protein.
OS Plasmodium yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5861;
RN (1)
RP SEQUENCE OF 379-2771 FROM N.A.
RC STRAIN=YM;
RX MEDLINE=97077455; PubMed=8920022;
Sinha K.A., Keen J.K., Ogun S.A., Holder A.A.;


```

QY 1065 PLRSNVLVHFGAGATNSMLPEPDEFELINSFARNVSNNSDNVAISGNASTISENOL-- 1122
DB 13187 -----TIVSELVAGVPRKEEALIPEDDVK-FRKPBERDAPEED-----SEIKLRPPQ 13222
QY 1123 ---DMNFDDQATIGOKIOEPASASANTVGGDDGLASAPETPTPTKESISSPKARLS 1179
DB 13233 ASKENDPEQALVTPK-AEPIIPEQIEDKAIIDEKKPKSKPKVKQPKQEPQIAKEEPEEF 13291
QY 1180 SASPRK-----SPKIGSP--VRIKKNCSIAGIEPIPKATHKKKSPQGNELIS-NHKV 1230
DB 13292 EVSAGEEALVDKPIETIEKPPDVKKKPKKPEAPSEVVVTEEPKPEEVEPIVEYKI 13351
QY 1231 RDGSISSSGSEHOHNSMVSVSQYDATSTVDEKNKDVONKREKOKONHNHNNH 1290
DB 13352 TTYVLEPEDAP--KEHOKVATIDFERQETTEEVI--EEKVYTRKKKPKRQOPEEEVNLK 13407
QY 1291 HHKOKTDPGVVDEIDPVDVGLQERKGLFVYLGIKNIMLPDINHKGRTTLONGVHCY 1350
DB 13408 EPKEEQIOPDVVSAEI-SLPIEE-----PEQKPRQYEVLEKITG----- 13445
QY 1351 TTPRYNMDDHVAIGKEPELVADSL--EFLITLKASYEK--KRTLVYVTEKKVYKSRN 1406
DB 13446 TTPPE-EPNDVQIAVKEKVKTKPVKKVKEDKIVVEAEKEEKQPEETIVEV-EKQEKKKS 13503
QY 1407 RLSLRFGSKDITTTTKFV--PTEVKOTWANKFAPDGSFACYIDLOQFDOITGK--AS 1461
DB 13504 EKPKSYEK--ISFTQSTIEEKPIEVAEE--APEET-----PKVYEKKVAKPEPSY 13549
QY 1462 QFDLNCFNEMETMSNGNOPMK-----RCKPKYIAQLEVKMLVPPSPREILPTSIR 1513
DB 13550 EFTLKEPDEEKVITVDQPEEAPVVEVFKKKPKKPEAVEAEFV--MTEPKIVETSVE 13606
QY 1514 SA 1515
DB 13607 TA 13608

RESULT 12
Q91704 PRELIMINARY; PRT: 6815 AA.
AC Q91704:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CG18242 protein.
GN CG18242.
OS Drosophila melanogaster (Fruit fly).
CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Goezanne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Boucek J., Brokstein P., Brothier P.,
RA Butlis J.M., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry K.C., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

```

```

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Melnikov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Paolel J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shie B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtens R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -i: SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: AE003473; AAC22226.1; -.
DR HSSP: P56276; 1TLK.
DR FlyBase: FBgn0035301; CG18242.
DR InterPro: IPR003962; F0111.repeat.
DR InterPro: IPR003961; FN_111.
DR InterPro: IPR003598; Iq_C2.
DR InterPro: IPR003600; Iq_1MC.
DR InterPro: IPR003006; Iq_1MC.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00047; Iq; 11.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00014; FNTYPE111.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00408; IGC2; 5.
DR SMART: SM00410; IG_1like; 6.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
KW Immunoglobulin domain; Repeat; SH3 domain.
SQ SEQUENCE 6815 AA; 779559 MW; A4E244001A4EBA01 CRC64;

Query Match 3.9%; Score 336; DB 5; Length 6815;
Best Local Similarity 16.9%; Pred. No. 8.5e-06;
Matches 286; Conservative 283; Mismatches 660; Indels 466; Gaps 69;

QY 41 TNSKPSLDPNSSDYTSSEODEGKEKKDPAFQTSFDRNFDLNSIDI-----QQ 92
DB 2740 TTTVLETPPDQOPVKKQRTKKIKKDEVEDVKVRIEEAPQREGSVLVYEDFVKP 2799
QY 93 TIOHQOQOPOOQOQOISQDNN-----LIDFSQTPMTSLDLTKONPYDKVNHNA 145
DB 2800 SSEKRRKKRPIKDKHTSVSEETPHEDEVLIESVPEDSPISDDLITVWVSVPREEENRV 2859
QY 146 PLYITSPKSKIMKATPASPCKVAFV-----TPEIHHYDNRV 187
DB 2860 NOIEDT-----KREKKKKRPPSKKILEENVPEOTVEKPLFALTTSDLEKPDVQESISIK 2916
QY 188 EEDDSQCKEDSVEPPLIIOHQMKDPSQFNYSDEDTNNAVPPRPLHTTKPTFAOLLKNN 247
DB 2917 EEOQHHTPEKKKSKISSEQKQSTBEQYELISVTHDLK-----EEKPFTVQYIQST 2972
QY 248 EY-----NSEEDA-LTDMKKLRNFNSNIS 270
DB 2973 NVEETKDDTGKVKHQQVTKRMLRRPAGECEIIEIVVRDQDEPAETIIVEAPE--PVN 3029
QY 271 IDEKTNVLVSPNTNNNSKVNVDMSHLOMLDQASKNKTN-ENIHNLSPFLAKPKNDIENP 329
DB 3030 QDEKPK---EPKKTTRKKKAKDDIHDIYQKLIELETPKTELEKYEKIEFPPIYKXPLDSP 3086
QY 330 LNSLTNADISLRSKSSQSSQSLQSLRND-----NRELVSVGSP----- 367

```

Db 3087 IDVLDESPKVEOKKDKSRSTKVPNEETPVQEOYAKVNVVEEAPDEPILVQILLEKPV 3146
Qy 368 -----KVNPGSL-----NDGIGSGEVESESL 392
Db 3147 EVDVKEVITEDGKVPQEKTTKRVLLKIGPEQOTFKITMIESEDNDSYTVYVDEPELAS 3206
Qy 393 PRDLSDKLETTKHDAPENHNEFIDAKSTNTKGLVSSDHLDSFDRSYHTEOSI 452
Db 3207 POST-----EEH--PEOSKEKLAPKPKTKVRK--VKKD-----L 3237
Qy 453 LNLNSASQOISLMALEKOKOTQEOGTOAEPPEETSFDNIKVQOEKPSNLEFVKV 512
Db 3238 SDVYKALLEETIPKVDLEKEVE-----MPEKPVKLTVSDISIPKPKD----- 3282
Qy 513 IKKBPVSA-TEIKAPKREFSSRLIKNE-----ETAEPAD 548
Db 3283 -KSQPIVLPPTKPKTKPTKPTEDTQOVPEDEPTETVDTTIDPELPTQAOED 3341
Qy 549 -----IPKKENANSIVEDTALLKALANDESDTTQNSTKMSIRPHIDSKLEDNS 603
Db 3342 TATAQITPSAOEEKSTQ-DTKDTIOKTVKHKTKRPTQKSVETSELPVAKDYQISIIH 3400
Qy 604 DG--DREDNDISRFKESDILNDVQTSDI--IGDKGNSSELTPTKTLAPPSDNNDK 658
Db 3401 EELVEEPEPEKILEVRVIDEAEVEESQPIVEEVEDEPQATEETVEDVKPS----- 3455
Qy 659 ENSKSLDEDPANNESLQOOLEVPHTKEDDSILANSSNIAPEELTVPEVANDYSFNDVT 718
Db 3456 -----KKKKVKKKTDH-----DELKKMLE-----QEIE 3481
Qy 719 KT-FDAYSFEESISREHETSKRINFISYHKOEKKHQIHKVPTQOIIASVOQYKNE 777
Db 3482 KTELEKTEKIEEDVPKPKKPEFALEPILIKIERKQK-----PKVTLL----- 3524
Qy 778 QESHVTSQVIVPAIOFKFK-----EVNV-----MSRRVSPMDLNVSOFLPE 824
Db 3525 -----DATVPRKTVKIKPKRKEKRAPEELTVQLPKRLARV--LVYEPAPRLIPK 3574
Qy 825 LSEDSGEKDLNFANYSNNTNRP--SFTPLSTKNVLSIDNDPVNEPPEPKSYAETRN 881
Db 3575 TTDTGALIKD--NGELSRNIEAEELIKFKPKTKT-KIKIDDELEKVELEKYEK----- 3626
Qy 882 ARRLSANKAANPAPLPPOQPSSTRNSNKRVSFRFVFPFELIRTSALAPCOMYNDI 941
Db 3627 -----ISSEEPREKTPYKKEAPK-----PEKQEDV 3655
Qy 942 FDDGAGSKPTIKAG-----MKTLPMDKDVRLNKAAGVQODEYINAKLVDOKEKK 996
Db 3656 KLLGKGKKKKKEEAPENYTLKNIPQKQ--EVEEVELKQKPEVELEVEQ--TKRKPD 3712
Qy 997 NSIYTDPEDKTEELQOTASINNATIDSIYGRPDSISTDM--LPVLSDELKRP-PTAL 1051
Db 3713 GEFVPER-----FEPESEFDRPEVVPDELEOIEHPELPEKVKPSKTKY 3755
Qy 1052 LSAORLFMEQEVNHLRSNLSVLYHPGAGATNSSMLPEPDELINSRANVSNNNSNVAIS 1111
Db 3756 KPKKKSSEPR-----TIVSEIVAGVKEEAPLEQDVK-FRKPREDADERT----- 3802
Qy 1112 GNASTISFNQL-----DMNFDDQATIGQIOEPASKANTVRCDDDLASAPETPRPT 1166
Db 3803 ---SEIKLRYRQASKDNPREQALVTPK-AEERIPQIEKALIDDEKPKKSKPKQV 3856
Qy 1167 KRESISSRAKLSSASPRK-----SPIKIGSP--VRVYKNGSIAGIEPIPKATHKPK 1218
Db 3859 KEQELKEREPEFYSVEEALVDKPIEIEKPKDVKKKPKAPSEVAVVEEPRK 3918
Qy 1219 SFQGNELIS-NKRVKDCGISPSGSGEHOQHNPSVSPQYTDATSTYVDEKKNQONKPRE 1277
Db 3919 EELVEEPEPEKILEVRLERDAP--KEHOYKVIDEFDERQETTEVI--EKKVUTRKKKP 3974
Qy 1278 KOKOKNNNNNNNNKOTDIPGVVDDIIPVGOERGLFRRVGLGINILNLDINTHK 1337
Db 3975 KPOOREPEVTLKEREQIOIPDVVSAET-SLPIEE-----PEOKPEQY 4017

Qy 1338 RFTLLDNGVHCYTPPEXNMDDHVAIGKEFELTVADSL--EFTLLTKASYEK--RGTL 1393
Db 4018 EVELKIQ-----TTPPE-EPNDVOLAAKEKVKTKPVKKVKEDKLVVVEAEERKQVETI 4071
Qy 1394 VEVTREKVKVSRNRLSRLEFSKDIITTTKVV--PTEYKDTWANKFADPGSFARCYIDLQ 1450-
Db 4072 VEV-EKEEKKKSEKPKSYERK--ISFQTSIEKPIEVAEE--APEET-----PK 4116
Qy 1451 QFEOITGK--ASOFDNCFMEWETMSNGNOPK-----RGKPYKIAQLEVKMLYVP 1500
Db 4117 VEKKAKEKFDSTYETFLKEIDEKVIYDDQPEEAPVEVFKKPKAPEVAEVEV--- 4173
Qy 1501 RSDPREILPTSIKSA 1515
Db 4174 MTEPKIVETSVETA 4188

RESULT 13

Q9NDS4
ID Q9NDS4 PRELIMINARY; PRT: 2678 AA.
AC Q9NDS4;
DT 01-OCT-2000 (Tremblrel, 15, Created)
DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)
DT 01-MAR-2002 (Tremblrel, 20, Last annotation update)
DE AMIB.
GN AMIB.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellidae; Dictyostellum.
OX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RX MEDLINE=20118106; PubMed=10651904;
RA Kon T., Adachi H., Sutoh K.;
RT "Amib, a novel gene regulated for the growth/differentiation transition
in Dictyostelium".
RL Genes Cells 5:43-55(2000).
DR EMBL; AB030033; BAB01489.1;
DR InterPro; IPR00194; ATPase_a/bcentre.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN.1.
SQ SEQUENCE 2678 AA: 298921 MW: 7082BFBFBEBETC7CA CRC64;

Query Match 3.98; Score 335; DB 5; Length 2678;

Best Local Similarity 18.18; Pred. No. 3e-06; Matches 288; Conservative 228; Mismatches 605; Indels 472; Gaps 65;

Qy 34 KLPNPTNSKPSLIDPNSSDPTYSQDOEGKEKKDTAFQTSFRNFDNSIDIOQT 93
Db 142 KPINSSNNS-----NITSSTISDLSLKNLKSPPSETTPPNTNNN--SNVYTDSP 192
Qy 94 IOHQOQPOQOQOLSQT--DNNLIDFESFOPTMTSLDITLTKONPTV-----DKVENHAPT 147
Db 193 PNATNKMSTSPKSLSPITSSNNNTTAAATTTTNTNNSNSPSPNTNNNNNNNSPS 252
Qy 148 ---YINTSPKNSIMKATPKASPKKVAFTYTNPRELHNYPDKRVEEDOSQCKEDSVEP-- 202
Db 253 SHNVVNSPP-----STSSRSRPTVASVTSN-----TSTIOPIS 285
Qy 203 PLIQHOKMDPSQVNSDSDTNASVPPRPLHTKPTFQQLLNKNNEVSEPEALJDMKLK 262
Db 286 PLINRQ--TSSHNIQO-----POQPYVNHQP-----LQPIQIYITTELM 323
Qy 263 RENFSNLSDKENVLYLSPNTNNNSKNVSDMD-SHLQMLDASKNKTENIHN--LSFA 318
Db 324 VFNISNTSL-----LDFSHLNVEGIVNLDNIGNNSVIYOYL 362
Qy 319 LKAPKNDIENLNLTNADIS-----LRSQSSQSSQSLQSLND--NRVLESVPSPKPV- 370
Db 363 KSLYNIIVE--YNNVSNDFIKIGNSFKSHILOKOKS-RNDYNNSTISGVSQSKRLF 419
Qy 371 -----NPGLSLNDIGKFSDEVEVESLLPRLSLDKLETTKEH 407

SQ SEQUENCE 3111 AA: 349635 MW: EDA9A5FD38115773 CRC64;
 Query Match 3.9%; Score 333.5; DB 5; Length 3111;
 Best Local Similarity 17.7%; Pred. No. 4.2e-06;
 Matches 274; Conservative 252; Mismatches 595; Indels 427; Gaps 65;
 27 STFNSTKLANPRNRSKSLDNPSSSDTYTSEDOEKKEKKOTAFOTSEDRNFDLN 86
 991 SFNNSLTQKMSSESSSKTPID-----LKDEDRPGSSSSK-----N 1026
 87 SLIDIOITQ-HOQOPOQ-----QOOLSDTNNLIDFSPOTPTSTLDTLTKONPTVD 138
 1021 QKDSEKTLKHSSEPTSTKETTGETTEASDNDKTEGKEPQRIKLPSTAE--LED 1084
 139 KYVENHAPTYI-NTSPNKSIMKATPKAPKVAFTVTPETIHVYPDNVEEDSOQKE 197
 1085 RENALEKKMSVQKSSPSKN--KKEPDEESK--STKEPE-----BPSEKKNEXT 1132
 198 DSEVPLIOHOKKDSQFNYSDEDTNNAVPT---PLHTTKPTFAQLLNKNVEYN--- 250
 1133 SGRQPTPIAKKDSQDQ---KKSETKENQSPTKNDEKVKVSPKSEEMIEKETSSNPKE 1189
 251 ---SEPEALTDKIK--RENENSLDEKYNLYLSPNNNSKNVSDMSHLONLQDASKKK 307
 1190 DSHESPAATNKKVEGNRELSSEKGDHKIKEKSEAPGKAGKETAEYKN--ANYKDSKKG 1247
 308 TNEINHNLSFALKAPKNDIENLNLTLNADISLRSOSSQSSLOSLRDNRYLESVPGSP 367
 1248 DSQ-----KNEAKTTSVSQTESD--LKPSKKNSTSKDAEQCK-----TP 1285
 368 KAVNPGSLNDIKGFSDEVE---SLPRDLSRDKLETTKEHDAPENHNEPDAKSTN 424
 1286 RKSPP-----STEELEKRFNALEKQMTNLLETKEPDDQK-----PAIK 1325
 425 TNGQOLVSSDHLDSFDRSYNHTBOSIINLNSASQSOISINALEKQOTQ---QEQT 481
 1326 SOSTAEVYTKQMSKSPDKIKEVNAV---EKQSRVVEVNAEKKKRKNVEAPRNKKG 1382
 482 QAEPEETSFSDNITKVQEPKS--NLEFVYTKIKPEVSTATEIKAPREFSSRLIRKN 539
 1383 DSQPEESQHKGNQRRASEPSTEDLEKRYETLKRMSKNOFSETVDEALERT---Q 1438
 540 EDEIAPRADIHKKENANSHEVDLALKKALNDESDTTONSK---MSIRPTISD 596
 1439 QVYISEAVY---EKKRPSTEDLESREALHGDKNVESKMDKEKHVDAVIEAIPIS- 1492
 597 WKLESDNDGREDNDISFEKSDILNDVQSOTDIIGDKY-GNSSEITTKTLAPRSON 655
 1493 ---PPPPPPPKERYLA---EPVLHQOQALIELOSKMRQSGEENLK---PSEIN 1541
 656 NDKENSKSLEDP-----ANNESLQOQLEVPHTKEDDSILANSSNTA 696
 1542 PQRRQKLLQRPMDTSEAPANTAYYRAANHEQOQRM-----VRRSDLP 1590
 697 PPEEL--TLRVVEANDYSSFNVTKTFRAYSSFEESLSNEHETDSKPTIFISIMIKQEQ 754
 1591 SRADLENRIQLEFOLYKFF-----YKORCASDSEVARS----- 1624
 755 KKHQIHKVPYTKOILASYOQYKNEOSRVTSDVKIPNALQFKKFEVNVMSRRVSPD-- 812
 1625 ---VKLPREDPSTSRQAKQOAEQOLQORVLALEKQISEMSKLLEMRERHSADDS 1680
 813 -----MDDLNVSOQLPELSEDSGCFDLNFA-----NYSNNTNRPRST 850
 1681 GSPRLSTETIDATGKELVRYTQNGILEEVDANHKRPINISINIKMMVNDSSKQKGS 1740
 851 PLSTNNVLSINDNPVNERPERPKSYAEIRNARRLSANKAANQAP-----LPQOQ 903
 1741 KPTEDLRLRLE-----QLEQOLLEERAKNGSIPPENEVLEEKPEKLE 1783
 904 PSSSTNSNKRVRPRVPTFEIRRTSALAPCDMYNDIRDDGAGSKPIKAEGKTKLPS 963
 1784 EKDSCKKQKKNCHQNVKQDEVEKTE---IPADRKIE-----PASKET-----KTLN 1829

964 MDKDDVK-RILNAKGV-----TQDEYI-----NAKLYDQKPKNSIYTPED 1005
 1830 VEAQGRRAVVDTEKSVKQGNVATDEKSYQDDQNVNVYDKAKDRKIDKXK-----SPAG 1884
 1006 RIEELQOTASIHNAITDSSITGRPDSISTDMPLPYLSDELKCPPTALLSADRLEMEQVHP 1065
 1885 KSEDTKQTS-----CKKE-----SEDIKQASEA----- 1908
 1066 LRSNSVLVHPGAGATNSMLPEPDELINSPARNVSNNSDVAISGNASTISPNOLDN 1125
 1909 -----PKAGAKETSTRGKPSSTLEKPTTESVUKETPPKKNLES----- 1950
 1126 PDDQATIGQKIOEOPASKSANTVRGDDGLASAPERTPTTKESISSPKALSS--ASP 1183
 1951 -----EKPKSKNEATK-----TEQKSKETPT--VAVSPRESKVSQOME 1990
 1184 KSPKIKGS-----PYAVIKNGSINAGIEPIKAT-----HKPK----- 1218
 1991 KKETIKDSSSKELPEKMYNSTDVGPMGPKTVVLLMDNEHRASKVRRLTRANTELED 2050
 1219 SFOG-----NEISNKHVRDQ---GISPSSGSEHOQNPMSVSPQYTPATSTVDEN-- 1268
 2051 LFOALEKQJLNDKRLVKSSEGLRIVDPKPSAQVEQTOAISDTLKEIETPSAKPEENP 2110
 1269 ---KDVQHKPREKOKQHNNRHHNNHNNKQT-DIPGVVDEIPVGOERKLPFRVLG 1323
 2111 KEAKEDKREPEREDFDMGPRMIVKHNILKRTVYLPSTKELSRFSRLQIKL----- 2164
 1324 IKNINLPDINTIKGRFTLLDNGVHCVTTPYENMDHNAVATGEFELYVADSLEFLITLK 1383
 2165 LEDVERKIDVEORL-----NEIERKIKLYSL-SHEKOLKLYLELCBGKGLD----- 2209
 1384 ASYEKPRGTLVEYTEKKVYSRNRRLSRLEGSKOITTTKFEVTEVYKDT 1431
 2210 --DDEVPRVETPTKEAETIATDR-SRSPGRKALATKSPYTSPPSKAT 2254
 RESULT 15
 Q17464 PRELIMINARY; PRT: 3147 AA.
 AC Q17464: Q22160;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE T04F3.1 protein.
 GN T04F3.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA White S.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.;
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berts M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Cratton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
 RA Smaiden N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RT Nature 368:32-38(1994).
 RL [3]
 RP SEQUENCE FROM N.A.

This Page Blank (uspto)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:20:59 ; Search time 32.8137 Seconds
(without alignments)
4875.027 Million cell updates/sec

Title: US-09-964-858-1
Perfect score: 8631
Sequence: 1 MNSTPSKLPIDKHSMLQ.....WVNLMLQ0000000SSSQ 1664

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1.*
2: p1r2.*
3: p1r3.*
4: p1r4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8631	100.0	1664	2 T18216	integrin-like prote
2	696.5	8.1	1320	2 S57113	BUN4 protein - yea
3	367	4.3	704	2 T50303	hypothetical prote
4	346.5	4.0	2401	2 T28676	rhophly protein -
5	342	4.0	2364	2 A56577	microtubule-associ
6	338.5	3.9	5327	2 T13564	microtubule-associ
7	337.5	3.9	1738	2 T14867	interaplin - slime
8	331.5	3.8	3147	2 T18674	hypothetical prote
9	329	3.8	6713	2 B89921	hypothetical prote
10	325	3.8	1650	2 T18444	hypothetical prote
11	324	3.8	1302	1 JC6009	surface-located me
12	320.5	3.7	1381	1 S45781	probable calcium-b
13	314.5	3.6	3724	2 T18427	hypothetical prote
14	314	3.6	2464	1 ORMSP1	microtubule-associ
15	313.5	3.6	3488	2 T14418	hypothetical prote
16	313	3.6	1658	2 S55101	hypothetical prote
17	312.5	3.6	1875	2 T32008	hypothetical prote
18	310	3.6	1875	2 S38173	myosin-like protei
19	309.5	3.6	1271	2 A45555	glutamate rich pro
20	304	3.5	2346	2 T13829	Tpt homology - fru
21	299.5	3.5	2269	2 T28677	rhophly protein -
22	297	3.4	2271	2 T90073	hypothetical prote
23	296	3.4	2139	2 T18296	myosin heavy chain
24	293.5	3.4	2481	2 D90011	FmB protein limpo
25	293	3.4	5170	2 T15348	hypothetical prote
26	292.5	3.4	1790	2 S67593	transporth protein
27	291	3.4	1744	2 JH0720	centromere protein
28	290.5	3.4	2663	1 S28261	breast cancer susc
29	290	3.4	3343	2 T42207	

30	285.5	3.3	1979	2 C71622	hypothetical prote
31	285.5	3.3	3869	2 A48205	All-1 protein +GFE
32	284.5	3.3	1727	2 T50073	myosin-like coiled
33	284.5	3.3	1957	2 T38077	hypothetical coile
34	284.5	3.3	4688	2 F82885	hypothetical prote
35	283	3.3	1364	2 T40839	hypothetical prote
36	282	3.3	3924	2 S37431	ankyrin 2, neurona
37	281.5	3.3	2722	2 T20532	hypothetical prote
38	281	3.3	1358	2 A29360	Sir4 protein - yea
39	280.5	3.2	1233	2 S36271	hypothetical prote
40	280.5	3.2	1435	1 BVBV11	guanine nucleotide
41	280	3.2	1139	1 E64234	cyathadene-acces
42	280	3.2	1466	2 A36426	SPA2 protein - yea
43	277.5	3.2	1189	2 S56852	hypothetical prote
44	277.5	3.2	1621	2 A82255	hypothetical prote
45	277.5	3.2	3225	2 I52300	giantlin - human

ALIGNMENTS

RESULT 1

T18216
Integrin-like protein alpha chain - yeast (Candida albicans)

C:Species: Candida albicans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000

C:Accession: T18216

R:Gale, C.; Finkel, D.; Tao, N.; Meinke, M.; McClellan, M.; Olson, J.; Kendrick, K.;

Proc. Natl. Acad. Sci. U.S.A. 93, 357-361, 1996

A:Title: Cloning and expression of a gene encoding an integrin-like protein in Candida

A:Reference number: 206510; MUID:96133936; PMID:8552638

A:Accession: T18216

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1664 <GAL>

A:Cross-references: EMBL:U35070; NID:q1144530; PID:q1144531; PIDN:AAA96019.1

A:Genetics:

A:Gene: alpha INT1

Query Match	100.0%	Score 8631	DB 2	Length 1664
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1664	Conservative	0	Mismatches	0
Indels	0	Gaps	0	
QY	1	MNSTPSKLPIDKHSMLQPOSSASIFNSPTKPLNFPRTNSKPSLDNSSDYTSQ	60	
DB	1	MNSTPSKLPIDKHSMLQPOSSASIFNSPTKPLNFPRTNSKPSLDNSSDYTSQ	60	
QY	61	DOEKGEKKKDTAFQTSFDRNFEDLNSIDIQOTIQHQOQOPOOQOQLSOTDNNLIDFSEF	120	
DB	61	DOEKGEKKKDTAFQTSFDRNFEDLNSIDIQOTIQHQOQOPOOQOQLSOTDNNLIDFSEF	120	
QY	121	QTPMSTLDLTQONPVYKVNENHAPTYINTSPNKSIMKATPKKPAKVAFTVTPETI	180	
DB	121	QTPMSTLDLTQONPVYKVNENHAPTYINTSPNKSIMKATPKKPAKVAFTVTPETI	180	
QY	181	HPDNVVEEDSOOKEDSEVPEPLIOHOKKDSQFVNSBEDTNAASVPPPLHTTPTPA	240	
DB	181	HPDNVVEEDSOOKEDSEVPEPLIOHOKKDSQFVNSBEDTNAASVPPPLHTTPTPA	240	
QY	241	QLLNKNEVNSEPEALTDKRLKRENFSLNLSDEKVNLYLSTNNNSKNVSDMSHLQNL	300	
DB	241	QLLNKNEVNSEPEALTDKRLKRENFSLNLSDEKVNLYLSTNNNSKNVSDMSHLQNL	300	
QY	301	QDASKKNTENIHNSFALKAPKNDIENPLNSLTNADISLRSSGSSQSLQSLRNDRYL	360	
DB	301	QDASKKNTENIHNSFALKAPKNDIENPLNSLTNADISLRSSGSSQSLQSLRNDRYL	360	
QY	361	ESVPSGPKVNVGLSLNDIKGSDVEVSLIPROLSRKLFTTKKHDAPEHNENFTDA	420	
DB	361	ESVPSGPKVNVGLSLNDIKGSDVEVSLIPROLSRKLFTTKKHDAPEHNENFTDA	420	
QY	421	KSTNTKGLLVSSDHLSPFRSYNHTQSLINLNSASQSLINLALEKOROTEOBO	480	
DB	421	KSTNTKGLLVSSDHLSPFRSYNHTQSLINLNSASQSLINLALEKOROTEOBO	480	


```

Db      1215 AKIDINLKVTKVNLNRNEDIQADNGGORNFTDWLVFNECQLVFDDGERITTFNAECSSNEK 1274
               :| | | | | : : : : | | | | : : : : | | | | : | |
Oy      1626 I-MICLOEIIYRNFRROPWYNLMLOOOOQQOQQOSS 1662
               ||| : | | | | | : : : : :
Db      1275 SDWMYKKDGEVELANVF-HQPWWKYCKCEKAEEKRTT 1311
               :| | | | | : : : : | | | | : : : : | | | | : | |

RESULT 3
T50303
Hypothetical protein SPAPYUG7.03c [imported] - fission yeast (Schizosaccharomyces pombe)
C.Species: Schizosaccharomyces pombe
C.Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C.Accession: T50303
R.McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Saunders, D.; Harris, D.
submitted to the EMBL Data Library, January 2000
A.Reference number: 225059
A.Accession: T50303
A.Status: preliminary; translated from GB/EMBL/DDDBJ
A.Molecule type: DNA
A.Residues: 1-704 <MCQ>
A.Cross-references: EMBL:AL136521; PIDB:CAB66312.1; GSPDB:GN00066; SPDB:SPAPYUG7_03c
A.Experimental source: strain 972ah(-); clone plasmid pYUG7
C.Genetics:
A.Gene: SPDB:SPAPYUG7_03c
A.Map position: 1

Query Match          4.3%; Score 367; DB 2; Length 704;
Best Local Similarity 21.0%; Pred. No. 3.5e-08;
Matches 189; Conservative 140; Mismatches 315; Indels 256; Gaps 38;

Oy      787 VKIPNAIOEKRR-----EVNVMRSRVVSPDMDDLNVSQLPELESDSGFKDLNFANYS 840
           :| | | | | : : : : | | | | : | | | : | |
Db      21 LRISPPISTDYECSDYASTIASISRESTMRNFRNSISTSPAESPDADGDSPFYD 80
           :| | | | | : : : : | | | | : | | | : | |
Oy      841 NNNTNPRSF-----TPILSTK-----NVLSNIDNPVVVEPPREPKSYAEIRAKRLSA 887
           : | | : | | : | | : | | : | | : | | : | |
Db      81 QTLNSSPSDDHQSLLPSTFEVRRTPTYSVNNETSSSTSVSE-----DVKNKENILSL 132
           : | | : | | : | | : | | : | | : | | : | |
Oy      888 N-----KAAPNOAPPLPQROPSSTRSNSNKRVSRFPVTFEIRRTSSALACMDYNIIF 942
           : | | : | | : | | : | | : | | : | | : | |
Db      133 NSCLIKLSDDEA-----SNKS-SKSTSPTRNSIKSNS----- 164
           : | | : | | : | | : | | : | | : | | : | |
Oy      943 DDFGAGSKPTIAEGMKLTPLSMDKDDVKRIKNAKGVTODEYINKLVDOKKKKSIYTD 1002
           : | | : | | : | | : | | : | | : | | : | |
Db      165 -NQHGRDIPi----- 177
           : | | : | | : | | : | | : | | : | | : | |
Oy      1003 PEDRYEELQOTASIHNATIDSSITYGRPDSISTDMLPYLSDELK-KRPALLSADRLEMEO 1061
           : | | : | | : | | : | | : | | : | | : | |
Db      178 -----PARSYCN SKL-----FNEDTLPAEFEEVISIPPKYL-----EL 210
           : | | : | | : | | : | | : | | : | | : | |
Oy      1062 EVNHPLRSNVLVHHPGAGATNSSMLPEDPFELINSPARVNSNNSNVAMISGMNSTISFNQ 1121
           : | | : | | : | | : | | : | | : | | : | |
Db      211 PTH-----SHNS-----DTSEFNS-----IVSSVSOMVGAGCINSIA--- 244
           : | | : | | : | | : | | : | | : | | : | |
Oy      1122 LDMNPDDQATIGOKIQEOPASKSANTRYRGDDGLASAREPTPTP-KKESSISKPAKLSS 1180
           : | | : | | : | | : | | : | | : | | : | |
Db      245 -SFGFSSEDSRPDIKTTPRLSFADENR-----ENCSTDIYRDSIHEYEPRLTS 293
           : | | : | | : | | : | | : | | : | | : | |
Oy      1181 ASBRKSPIKIGSPRVYIKKNGSIAGEPIPKATHKRPKKSFOGNETISMHKVDRGISPSERG 1240
           : | | : | | : | | : | | : | | : | | : | |
Db      294 -----STSLDSP-NHYLDEN--APIRLLRKVVASLP----- 320
           : | | : | | : | | : | | : | | : | | : | |
Oy      1241 SEHOCHNPMVAVPSQYDTATST-VREDKKVDONHPRKOKOHNNHHNNHHNNKOKTDLP 1299
           : | | : | | : | | : | | : | | : | | : | |
Db      321 -----DRPFTVLVSFDALTRTYLLRONSKVAAHTSOEQMOTSRVAUVNSCYMPES-LS 373
           : | | : | | : | | : | | : | | : | | : | |
Oy      1300 GVAVDEIDPVYGLOERKGLFRRVLGIKINILPDIINHKGFTFLTDNGVNCVUTPRENMDD 1359
           : | | : | | : | | : | | : | | : | | : | |
Db      374 RNLISSSQGTG--GGGRGLFVRMLEIRNLTIPLASGMTTRREFYTI-SGNH-IDVP-WNNLN 428
           : | | : | | : | | : | | : | | : | | : | |
Oy      1360 HNAVIGKEFELTVADSLFEILTLLKASYEKPRGCTLEVTEKKVYKSRNRLSLRFSGKDITTT 1419
           : | | : | | : | | : | | : | | : | | : | |

```

```

Db 429 STTKIENEYFDESISSIVCTLRAAYDPRK-----VTRSTIGKVEST-----N 473
Qy 1420 TTFEVPTE-VKDTANKEFAPRGSEFARCYIDLOQFEDQITGKASQFPLCNEN--ETMSN 1476
Db 474 KRKSMTTDPAVSEALHGFSESDGTGEVITINDSVSRALRGCSMYLPIMNKWTVPDPAK 533
Qy 1477 GNPMPKRGKRPKIAOLEVKMLVYPRSDPREI--LPTSTSAVESINELNNEQNNYFEGYL 1534
Db 534 DVKPL-----FRKVELELHVFFLP-ALPVSLEKELPASTISAMYDLAEMDRFLLCDGYL 588
Qy 1535 HOEGDCPIFFKRPFKLMTGSLLAHSEISHKTRAKINSKVVDLI---VYDKENIDRSN 1590
Db 589 CQGGGDCPYRRRRYEQJIGSKLVAFQGFSKVRATIDISEAHNIYVDNHNYSDEBELE--- 645
Qy 1591 HRNPSVLLLDHAKKIFANSELIDFCAPRNHEKMIWIONLOEITIRNFRPQWYNLML 1650
Db 646 ----GYLFEESGPRRIIFSNDYIDFVAETVGEKDEKMSTLRHLLQCSKVHKNMTKSF 700
RESULT 4
rhotry protein - plasmodium yoellii (fragment)
C:Species: Plasmodium yoellii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28676: A45521
C:Stnpha, K.A.: Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: 220507; MUID:97077455; PMID:8920022
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: EMBL:U036927; NID:91041784; PID:91041785; PIDN:AA841263.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoellii rhotry protein. Multiple co
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: GB:M34281

Query Match 4.0%; Score 346.5; DB 2; Length 2401;
Best Local Similarity 18.9%; Pred. No. 1.3e-06;
Matches 360; Conservative 327; Mismatches 736; Indels 477; Gaps 94;

Qy 12 DKSHILOLOQSSASIFNSPTKPLNFPRTNSKPSIDPSSSDTYNSDEQEGKEE--K 69
Db 603 DKTELTETGTLNLNHNESNNKELTYFYDLKANCGKNKNLTKQNEKRAVEDIKK 662
Qy 70 KD-----TAFQTSFDRNFDLNSIDIOQTIOHQOOOPQOOOQLOSTDNNLIDE 117
Db 663 KKNVDINKIVSNIETIYTSIININEDTEN--EIGKSIELNTRKYLEKRVANVNLNLEIKE 720
Qy 118 ----ESFQT-----PMTSTL--DLTKQNPYDVK-----NENHAPTY--- 148
Db 721 KLKDYDQDFQEGEKENIKRYPDENKIKINDIDTLNQKIDKSIETLTIKKNSHNHIDEIKQI 780
Qy 149 --INTSPNKSIMKATPRASPKKVAFTV---TNPEIHYPNPRVVEEDSOOKEDSVER 202
Db 781 DKLRKYVKNKTFMNE--DPKEIKEIKIENIVEIKDKRNKIYEDIKLNEISKIENDKTSLE- 838
Qy 203 PLIOHQMKDPSQFVYS-----DEDTNMSVPTPTPLHTHTKTPPA-----QLL 243
Db 839 -----KLNKINLSGKSLGNLFLOQIDEEKKAE-----HTTKAMEAYIDDLDNIK 884
Qy 244 KKNNEVNSEPEALDMLK---KRENFNSLSDLEKYNLYLPTNNNNNSKNVSDM--DSHLON 299
Db 885 KKSQEIKEKNININDIKMDIKEMKALNISHD--KYIY-HTSKNHEDEKSDLRKNSLKI 942
Qy 300 LODASKKTKYENIHNLSPALKAPKNDIENPLNSLTNADISLSSGSSQSLQSLRDNRY 359

```

Db	943	IQDSEESIYINDIK-----ELEKNVLESNN-----NTDIN-----QYLSKINITYNI	986
QY	360	LESVGSPPKRVNPGLSINDIKGFSDEVEYS--LPRDLSRDK--LETTYKEHADBP--	412
Db	987	L-----KLNKIKRIIDKVEYTDIEIKNNKKINIAELNSEKIIITQOLKENSLSKECOS	1038
QY	413	-----NNENITDAKSTN--TKGQGLLYSSDDHLDSPFRSTNTHQOSI--LNLN---SASOS	462
Db	1039	KIKSTIDNDVYSECIKNTILKTYIYENKNNINNTYKMAEYONVSLFNFNIEAOWKS	1098
QY	463	QISNALEKOROFODROT--QAAEPBEETS--FSONI--KVQOPKSNLEFV-----KTYI	513
Db	1099	OYILN--IKKNGNTNDYINIKELKEIKKKSNNYKDDAGANTOEIKKNKELFEKYQOEYTV	1157
QY	514	KKEPVASATEIK---APKREFSSRLR--IKN--EBDEIAEPADIPHKRENE--ANSHVEDTAL	567
Db	1158	LKNYVAVELKKNFKEDKKNVSEQIIEIKIDAHTPTISQADKSEKKMNEIKNQIINIEDEV	1217
QY	568	LKKALNDESDPTQNSTKMSIRF-----HIDSMKLEDSNDGDRE-----DNDDISR	615
Db	1218	AKNNKSKALIDLOLSEVPKIFELKIDKLRKSDCLCETKDIETKISNSLIDPOTKRL	1277
QY	616	FEKSDILNDVYQSOTDIIGDYKNGSSSEITTKTLAPRSDNNKENS--KSLDDPANNESL	673
Db	1278	IENKNILNTLEKLLSEKKNQKIIDQ--KKEL-----DEVASKIKNISVNNQKK	1322
QY	674	QOOLEVHTREDDSIANSSSNIAPPEELLPVV-----EANDYSFNDYKTFDAYSS	726
Db	1328	NYEIGIYE--KINEIAKANKQIESQOLKIPIITKILNISPFKANDLEGI--DTNKNLGKYNT	1385
QY	727	FEELSLEHETDSPINFISIMHKOKKKHQIHKPTQIILASVOQYKKNDEGSVYTDK	786
Db	1386	--EMNNIYEEFIKSYDLIT-----HYLETYSKEPI--TYEQIINK--RITQON	1427
QY	787	VKIPNALQOFKKF-----EVNYSRRVY--SPDMODLVNSQFLPELSDSGFKOLNAN	838
Db	1428	ELLNINIKNVKAKSYLDDIEANEFDRIVHFKKLNDV--DKFTNEYSK--VAKGDN	1482
QY	839	YSNNTNRPSEFTPLSTKRVLSINIDNDPNVVEPERKSVAEIRNARRLSANKAPNOAPPL	898
Db	1483	ISNSINNWKST--DENLLNITLNOTK-----EMYANIVSKYKYXYEAEENIFINI	1532
QY	899	PROQOPSTSNKSKRVRVPFELRKRSSALACDMYNDIOPDFGSGSKPTIKACGM	958
Db	1533	P-----KLANSLNIOI-----KSSSGI-----DLFKINIAIILPYLDOSK	1568
QY	959	KTL---PSMKODVKRILNKKGVTODEYINAKLVQGRKKSIVYDPREDVEELQOPAS	1015
Db	1569	DTLFIPIPSPEK-----TSEYI--TKISQSYTLTDLKRSEOLOKKKQOQALN	1613
QY	1016	I--HNATIDSSIYGRPDISTDM,PLYLDELKPKPALLSADRLPMEQ--EYHPLRSNSV	1071
Db	1614	LIFENRLLHDKVQ-----ATNELKOTLSDLKAKKKKQOILNKYKLLKKNELKMSLCSNS	1666
QY	1072	LVPBGAGATNSSMLPEPDELINSPARNVSNNSDNVAISGNASTISFNQMDMNFDOAT	1131
Db	1667	-----QNVDTLESSKVDKIKERSNNYEKEKEMGI-----NFDVKA--	1703
QY	1132	IGKQIQOPASKANNTYVRGDDGLASAPETRPRTPKESISSPKAKLS-----SASPRK	1185
Db	1704	-----MEQOFNNDIKDIEKLENNYKHSKEKDYNESEENNNILOSKKKLELTNAFNAELIK	1759
QY	1186	SPIKIGSPVRKIKNGSIAEI-----EPKIPAKNPKK	1218
Db	1760	IEDI-----IEKNGLINKLIETRKDCMLFTYKTLVEFLKIKTIDTYKFTISATKFSKE	1813
QY	1219	SFOGNETSNHKVRDQGISPSGSEHOHNPVNSVPSQYT-----DATSTVPD--	1266
Db	1814	FLAKITDNTSMLNDIITLOTCTYDLNQIKKHVASVADATNDNNMLIEKEKATATGINNL	1873
QY	1267	-----ENKDVQ-----HKPREKOKOKHHNRRHH-----HHHKOKT	1296

Db 1874 TELFTIDSNKIDADGLHNKKIOIIFYNSELSHKSIDSIKOLYKKHMAFKLLNIGHINKKYF 1933
Qy 1297 DIRGVVD-----DEIPDVG--LOERGLKPRVVGIGKININLPDINTHKG---- 1337
Db 1934 DISKEFDNLOEOSELATNLDLKEIGOKISDKRKQFLHAF--SETPLEPNLTJKEIYH 1991
Qy 1338 ---REFTLIDNGVHCVTTPPEYNNMDHNVAGKEFELTVADSEFILTTLKASYEKPRGTL- 1393
Db 1992 DIVAKKNQIDE-IEMIT-----NEENENITLYMDIITKLKKKQOSIINFPYTTENDSNIITK 2046
Qy 1394 -----VEVTEKKVYKSRNLSRL---FGSKDITITTTTKFVPEYKDTW 1432
Db 2047 OHIDONNENDVSKIKESLETTIOSFQKILNKLNIEIKAOFYDNNNNINISTISQVDVY- 2105
Qy 1433 ANKFPDGSFACFYIDLOQFEQITGKASQFDLNCFNEMETMNSGNQPKRCKPYK--IA 1490
Db 2106 KKHISKDLTENLQIOKSLSDI--KKSTYDI-----RSBQITKYVNPPIHDVYE 2153
Qy 1491 QLEVMLYVPRSDPREILPTISRSAYESINELNNEONNYFEGYLHOEGGDCPIFKKRFK 1550
Db 2154 QOTKTIQNNPKD-----EIDDLIQELVYNNKESLKLPTIINKNDVTEPII--SRIDK 2205
Y 1551 LMGTSLAHSEISHKTRAKINLSKVVDLIYDKENIDR--SNHRNFSVDLLLDHAFKIK 1607
Db 2206 VIN--LIKSEYNNNDVSNVAKKLE--EDANNIIRDLDTSHNNLND--LIQKNFKI- 2256
Qy 1608 FANGELIDPCAPNKHMKIWIIONLOEIIYRNFRQPPVYN 1647
Db 2257 -----IDDLKNNKOEIEN--RNNIQIINREOEITQTEHYN 2289

RESULT 5

microtubule-associated protein MAP 1B - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Feb-1997
C:Accession: A56577
R:Zauner, W.; Kratz, J.; Staunton, J.; Felick, P.; Witche, G.
Eur. J. Cell Biol. 57, 66-74, 1992
A:Title: Identification of two distinct microtubule binding domains on recombinant rat M
A:Reference number: A56577; MUID:92347374; PMID:1639092
A:Accession: A56577
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2364 <ZAU>
A:Cross-references: GB:X60550
A:Experimental source: brain
A>Note: nucleotide sequence not given; conceptual translation not complete
C:Superfamily: microtubule-associated protein MAP1B

Query Match 4.0%; Score 342; DB 2; Length 2364;
Best local similarity 17.8%; Pred. No. 1.9e-06;
Matches 349; Conservative 313; Mismatches 695; Indels 608; Gaps 92;
Qy 2 NSTPKLLP-IDKSHLDL--OPOS-----SASIFNSPTKPLNF--PRTSKPSLDPNSSS 53
Db 396 NSTQYNIIEGLEKLNHDLKPLATQKDLGOVSTPRPKQYKILKORADSRSLRP--AT 453
Qy 54 DTYTSEDOEKGEKKKDDTAFTQSFDRNFDLNSIDIQOTIOHQOQOPOQOQOOLSOTDNN 113
Db 454 KPLSSKSVRKESKEAPEKTAQOYEKT-----PKVESKEKVIYKKDKPKG----- 499
Qy 114 LIDESFQTPMTSTLDLTKONPTVDKVNENHAPTYINTSPKNS-IMKATTPKASPKKAVF 172
Db 500 -----VESKPS-----VTEKEVPSKEEOSPVAEVAEAKATSKSPKVTKD 539
Qy 173 TVTNEPIHHYPDNRVEEDOSO--OKEDSV-----EPPLIOHQWQKPSQFNVSDE--DTN 223
Db 540 KYVKEIKITKPKKEKPKKAKKEDKTPPLAKDKPKKPKKEAKKIKKEKEKELK 599
Qy 224 ASVPPTPLHTTKPTFAOLLNKNNEVNSPEALJTMKILRENFSNLSLDEKYNLYLSPTN 283
Db 600 KEVKKETPLKDAK-----KEVKKDEKKEVKKKEKPKKEIKKISKDIKKSTPLSDTK 651

Qy 284 NNNS--KNVSDMSHLQ-----NLQDASK-----AKTNENIIN-----LSF 317
Db 652 KPAALKPKYAKKEEPPKPKPIAGKILDKGKVKIKKEGTTTEAATPAVGTAAVAAGA 711
Qy 318 ALKAPKNDIENPLNLTNDISLRSSGSS-----OSSLOSLRNDNR 358
Db 712 AASGPAKLE-----AFRSLMSSPEDLTDFEELKAEIDVAADIRPQLELIDEEK 763
Qy 359 VLESVP-----GSPKKVNPGLSLNDGIKGFSDVEVESLP-----RDLNR 398
Db 764 LKTEPEGEAVYQKETEVKGSAESPEDEGITTEG--EGCEQTPPELEPEKOGVDIIR 822
Qy 399 -----OKLETTKHDAPENHNENFIADKSTNTNKGQLVSSDDHDS 440
Db 823 FEDEGAFGEESSEAGDYEEKAE--TEBAEEPEDEGDNVSGSASHSP-----TEDEIAK 876
Qy 441 FRSYNNHTQOSILNLTNSASQOISLN-ALEKQROTQEOQOAAPEEETSFSDNIKYK 499
Db 877 LEADV-HIKEKRSVASGDDRAEDMDALEK-----GEABQSEBEBEEDKADAREED 931
Qy 500 QEP-KSNLE-FYKVTIKK-----EPVSATEIKAPKREFSSRIILIRKNEDE 542
Db 932 HEPDKTEADYVAAVVDKAAEAGVTEDQYDFLCTPAKQPGVQSPSREPASSI-----HDE 986
Qy 543 IAEPAIHPKKEANEASHVEDDALKALNDEESDTQNSTKMSIRHIID----- 594
Db 987 -----TLPGSSESEATAASOENRE-----DDPEEFTATSGYQSTIEISSEPTPMDEMS 1035
Qy 595 --SDMKLESDNDGREDND-----DISRFKES-----DILDVQSOTSIIIGD 634
Db 1036 TPRDVTIDETNNETSPSPQEFYNITIKYESSLSQESKRYVASFNGLSDGSTDATDGR 1095
Qy 635 KYGNSSEITTTKTLPADPRSDNNDKENSLSLEDPANNEISLQ-----QOLEVPHTKEDDSIIA 690
Db 1096 DVNASAS-----TISPPSSMEEDKFKSKALRDAYREEDTVKGAELDKIDVDERLSPA 1150
Qy 691 NSNNIAP-----PEELTLPYVE-----ANDYSSFNQDTKTDAVSPFESISREH--E 736
Db 1151 KPSLSLSPSPSPLEKT--PLGERSVNSFLPNETIKASAEBAATAVSPGYQAQVAEHCAS 1209
Qy 737 TDSKPINFISIMHKOEKOKKHQIHVKPTKOIYAS-----YQOYKKEOES-----RYSOK 786
Db 1210 PEKKTLEVVS-----PSQSYGSAHTPTTYQSPTEPKSHLPTVEYENA 1253
Qy 787 VKIPNALIQKKEKEVNVMSRRVYSPDMD-----LNVSQLPELSDSGFKDL 834
Db 1254 QAVPVAFEEFTEAKDEN--ERSSTISP--MDEPVDPSEPIEKVLSPRLSPGISESAYED- 1309
Qy 835 NFPAVYNNNNRPRSTPLSTKYNLSMIDNDPNVYEPPEKSVAEIINARLSANKAAPNQ 894
Db 1310 -FLISADKALGRSESPEFGKNGKQGFSDKESVSULTSDLVODKOEKRAK----- 1360
Qy 895 APPLPQROPQSSSTRNSNKRVSFRVPTFEIRTSALAPCDMYNDIFPD--FGAGSKPT 952
Db 1361 FPIPKEDFSPREKASDA-----EIMSSQALNA-----LDERKILGGGSGPT 1400
Qy 953 IKAEGKTLPSMD-----KDDVKRILNAKKGVTOXEYINAKLVQDKPKKNSIVTDPPD 1005
Db 1401 -----QVDVSQFGSFKEDTK--MSISEGTVS DK--SAPRYDE-----GAED 1437
Qy 1006 KYEELQOTASINATIDSSISYGRPDSISTDMPLYLDELKRP-----PTALLSADR 1056
Db 1438 TYSHMEGVASVSTASVATSSF--PEPTTDDVPSLSLAEVGSPSTEVSDLSVSYQOTPT 1495
Qy 1057 LFMEOGVHPLRSNSVLYHVGAGAGATNSMLPEPDFELINSAPRN--VSNNSDVAVLSGNA 1114
Db 1496 TFOETMSPSKER-----CPRPMSTSPDPSPKTAKSRTPVQDHRSEDSMS----- 1542
Qy 1115 STISFNO-----LDKNFD-----DQATIG----- 1133
Db 1543 --IEFGQESPEHSLAMDPSRQSPDHPPTVGAGMLHTENGPTVEVDYSPSDIQDSSLSHKIP 1600

QY 1134 -----QKIQEQPASKSANTVAGDDDDGLASAPETPTPKKESIS-- 1172
Db 1601 PTEEPSYODNDLSELISVQVEASPTSSAH-----PSQIASLEDETLSDV 1649
QY 1173 -----SKPAKLSASPRKPIKIGSPVRIKNGSIAGIE 1207
Db 1650 VPRPDMSLYASLASKVQSLSEKLSIPKSDISPLTPRES-----SPYSPGSDSTSGAK 1704
QY 1208 -----PIPKATHKP---KKSFOGNEISNHKVRDGSIPSS-GSEHQOHNPSMV 1251
Db 1705 ESTAAVQYSSSPIDAAAAPRYGFRSSMLFTMOHNLALSRLDTSSVEKNDGKTPGDF 1764
QY 1252 SVPSQYATATVPDENKDYOKHPRKKOKKNNHHNNHHNNHKKOKDIEGVND----- 1304
Db 1765 NYAVOKPESTESPDPE-EDYDYESHEKTIQAHVQGYTEKERT-IKSPGDSGYSETI 1822
QY 1305 -----EIPDVG-----LOERKGLFPRVLGKININLPDIN-----THKGRFTL- 1341
Db 1823 EKTTPEDGGYSCITEKTTPTREGGYSYEI-SEKTTTPREVGYTEKTERSRLLD 1881
QY 1342 TLNDGVHCVTTPPEYMDHNAIGKEFELTVADSLFELITLKASYEKPRGTLVEYTEKV 1401
Db 1882 DISNVDYDTEODGHTLGCCSYETTERKIT-----SPSESYSYE-----TTTKT 1927
QY 1402 VKSRNRRLSRLFGSKDIITTTKFPVTEVKDTYANKFAPD---GSFAP-----CYIDLQDFE 1453
Db 1928 TRSPDTSAVCYETMKIKITTPQASTYSETSRCTTPREKSSSEKROVDLCLVASCERK 1987
QY 1454 DQITGKASQF---DLNCFNEMETNSNGNQPKMR--GKPYKIAOLEVKMLVYPRSDPR-- 1505
Db 1988 HPKTELSPSFINPPLJEMFAGEPTEESERPLTQSGAP-----PPSGGKQO 2034
QY 1506 -----ELIPTSIRSAVESINELNBNQNNVFEGYLHQEGDCCIFKKRPFKLMGTLLAHS 1560
Db 2035 GRQCEPTPTSYSEAPQOTDSD-----VPPETECP-----SITADA 2072
QY 1561 EISHKTRAK-INLSKYVDLIYDKENI---DRSNHRNFSVDYLLD 1601
Db 2073 NLDSESETIPTDQTYTKHMDPRPAMQDRSPSRHPDVSMD 2117
RESULT 6
113564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
Accession: J113564
S:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: 217689
A:Accession: J113564
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
C:Genetics:
A:Cross-references: EMBL:AL031128; PDB:CAA20006.1
A:Cross-references: FlyBase:FBgn0025392
A:Initrions: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A:Note: EG:49E4.1
C:Superfamily: Drosophila 576k microtubule-associated protein homolog
Query Match 3.98: Score 338.5; DB 2; Length 5327;
Best Local Similarity 19.48: Pred. No. 8e-06;
Matches 345: Conservative 292; Mismatches 713; Indels 425; Gaps 77;
QY 5 PSKLLPIDKSHLQLOP-QSSSASIFNPTKPLN-FPRT--NSKPSLDNPSSOTYTSQ 60
Db 2555 PTLSSPVQV-AHASVQPELSKVDEKTAASSPIDAPKSLIGSPAERESPAP--SAK 2610
QY 61 DOEKGEKKKDTAFQTSDFRNFDDNSI-DIQOTIQHQOQPOOQOQSLQDNNLIDFES 119

Db 2611 DAAESVEKSKDASRPSPVSESTRKADSTGDISPSESVLEGPKDDVEKSK-ESSRPPSVS 2669
QY 120 FQTPMTSTLDTLQKONPTVDKVNENHAPTYINTSPMKSIK-----KATPKASPKV 170
Db 2670 ASITGDSFTFDVSRPASVYESVKDEHDKA---ESRRESIAKVSVIDEAKGSKSSSQ-- 2724
QY 171 AFTVTPPEIHHYPDNRVEEDOSQOKEDSVPEPLIQHQKQDPSQPNYSDEDTNNAVPTPT 230
Db 2725 -----DSQDEKSTYLAKEASRRRESVSSKDAEKGESRPSVIAAGEPVP 2771
QY 231 -----PL---HTTKP---TFQNLKNNEVSEDEALTD---MKLKRNFN-----LS 270
Db 2772 RESKSPLOSQDTSRGSVYESVTADEKESQDSRRRESVAESYKADTKDKGQKQESRPS 2831
QY 271 LDEKVNLYLPTNNNSKNVSDMDSH--LQNLQD-----ASKNKTNENIHLSPALKAP 322
Db 2832 VDE---LLKDDDEKQESRRQSTGSHKAMSTMGDSPPMDKADKSKPEPSVASINHE 2888
QY 323 K-NDIENPLNLTNADISLRSSGSSQSL-QSLRDNRYLESVPGSPKKVNDGLSINDGI 380
Db 2889 NTKDEESPL-----GSRDVSVAESIKSDITKGEKSPLSKREVSHPESVGS 2935
QY 381 KGFSD-----VESILPDLSDKLET--TREHDAPEHNNENFTDAKSTNTKQQLVS 433
Db 2936 KDEKAEKRESVAESVYK-ESSKDATSAPPSKEHSRPESVLSGLDDEGDKTTSRVSVAD 2994
QY 434 S--DDHLDSFDRSYNHTQSLINLNSASQSQ-----ISLNALEKQROTQEPOTQA 484
Db 2995 SIKDEKSLVSOEASRPSEASESLKDAAPASQETSRRPSVETSVYDGGKSPVASKASPA 3054
QY 485 EEEETSTSDNKKVQOEKSNLEFVYKVIK--KEVVSATEIKAPRRESSRLRKNE-- 540
Db 3055 SVAEAKDSAEKSKQORESLPQSKAGSIKDEKSP-ASKDEAEKSKESRRESVAEOPF 3113
QY 541 ---DEIAPADIIHPKKEKEDANSHVEDTALLKALN-----DDESDTQNTKMSI 589
Db 3114 LYSKESVSRPASVAESVKDEAEKSKESPLMSKEASRPASVAGSVYDEAEKSKESRRRSV 3173
QY 590 RPHIDSDMK-----LEDSNDGREDNDISREKSDIIN-DVYSQTSDI--IGD 634
Db 3174 AEKSPLPSEKASRPASVAESVYKDEADKSKESRRRESGAEKSPASKASRPASVAESIKD 3233
QY 635 KYGNSSE-----ITTKLAPRSDNNKDKENSKSLIEDANNESLQOQLEVHTKEDS-- 687
Db 3234 EAEKSKESRRRESVAESKSPLSKESRPTSVAKSVK-----EAEKSKESRRSD 3282
QY 688 ILANSNIAPPEELTLPVVEANDYSFNDYTKTFPAYSSFEESLSREHETQSKPTNF-S 746
Db 3283 SVAEKSPILA-SKEASRPASVAESVODEAEKSKESRRRESVAEKSPYLAKEASRPASVAES 3341
QY 747 IWHQEKQKKH-QIHKVPYTKQIIAS-----YQYKNDQE----- 779
Db 3342 IKDEAEKSKESRRRESVAESKSPASKASRPTSVAESVYKDEAEKSKESRRSDVAEKSP 3401
QY 780 -SRVTSQDKVXKIPNLIQ--FKKFEKVNVSRRVYSPDMODLVNSQFLPLSDSGFKDLNF 836
Db 3402 ASKEASRPASVAESVODEAEKSK--ESRRESVAEKSPASKASRPASVAESVKD--D 3456
QY 837 ANYSNNTNRRPSF---TPLTSTKNVLSNIDNPVNEVEPEPKSYA-----ELRNARLSAN 888
Db 3457 AEKSKESRRRESVAEKSPASK-----EASRPASVAESVYDEAEKSKESRR 3503
QY 889 KAAPNAP-PLPQORPSSTRSNSNKRVRFRVPTFEIIRTS-----SALACDMYNDIFD 943
Db 3504 ESVAEKSPLPSEKASRPPTSVAESVYKDEAEKSK--ESRRESVAEKSLASKE----- 3553
QY 944 DEGAGSKPTTAAEGMKTLPSSDKDDVKRIILNAKKGYODEIYNNKLVDQKKKSIYTOP 1003
Db 3554 ---ASRPASVAESVKDEAEKSKESRR-----ESVAEKSPASKASRPASVAESVKDE 3604
QY 1004 EDRYEELQOOTASI--HNATIDSSIYGRPDIS-----TDMIPYLSDE 1043
Db 3605 AEKSKESRRRESVAEKSPLPSEKASRPPTSVAESVYKDEADKSKESRRRESGAEKSPASME 3664


```

OY 1094 INSPPRANNSNDVNAISGNAISTIFSNQMDNFDOATGKIOE-----OPASKSANTYR 1149
Db 1242 OSIOPENDEKEKQLOSEKDEKLOSIQONLNQLDNDENDEKKAQOSEKDEKLOSIQODLNQK 1301
OY 1150 GDDDLGLASAPETPRTPPKKESISSKPAKLSSASPKSPKIKGSPVAVIKKNGSIAGIEPI 1209
Db 1302 QENQ-----EKQKQSEKDEKLOSIQODLNQNDQO-----IKKNKLEKE-- 1343
OY 1210 PKATIKPKKSPGNGEISNNIKYVDGIGSPSSGSEHQHNPSWVSFSQITDAISTVPDENK 1269
Db 1344 -EQLKLQODFDNDOOQOLKOLEKISEKENQLOOLKQENELNOLNQOOS-----NE 1395
OY 1270 DVONKPREKOKNNHNNHHNNKOKTDIPGVVDEIPRV-GLQERKGLFFVVLGIKNN 1328
Db 1396 ITQOLKQDLKQKQOQOQOQOENNEKE-----IERLIOEIQOLKQOQOQOQOSELSSKEIK 1448
1339 LPDINTHKGRFTLLDNGVHCVTTPPEYNDMDHNVAIGKEFELTVADSEFILTLLASYEK 1388
1449 IQ-----TTQEEFDLSHN-----RSMDOQLHQLOQOELQDLKQSFSD 1486
OY 1389 PPGTLEVEYTEKVVKSRNRL-----SRLEGSK-DITTTKVPVPEVKDTMANKFAPD-- 1439
Db 1487 Q-----DHQFKKVIDERYMQLOLEOSTLSNNQDLQLEKLEKPLELDSNEKOKTIDDL 1541
OY 1440 GSFARCYIDLOOFEOITGKASQF-----DLNCFEMETMSGNGPMKRGKPYKTA 1490
Db 1542 SNISNQLQISLQNDKDLISERNNSITLTESRITQQLSLDEKQNLKLOLQOQOQOQO----- 1596
OY 1491 QLEVMKLVYPRSDPREI-----LPTSIRSAVESINELNEONNYFEGYLNQEGGDCPTFKK 1546
Db 1597 QQPPTASSPSSPSLSTPTPKPQRPQIQEIDRLVNVIVARNQDLIKN-----KT 1649
OY 1547 RFFKLMTGSLAHSEISHKTRAKINKSVVDLIYDKEN 1585
Db 1650 KEYKLENGDYIVNSII--YRLSLDDNDSDLIAQEYEN 1685

RESULT 8
T18674
hypochemical protein T04F3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T18674; T24464
R:White, S.
submitted to the EMBL Data Library, June 1996
Reference number: Z19004
Accession: T18674
Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3147 <WIL>
A:Cross-References: EMBL:Z74026; PIDN:CAA98419.1; GSPDB:GN00023; CESP:T04F3.1
A:Experimental source: clone B0240
R:Kershaw, J.
submitted to the EMBL Data Library, May 1996
Reference number: Z19894
Accession: T24464
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3147 <WIL>
A:Cross-References: EMBL:Z72513; PIDN:CAA96672.1; GSPDB:GN00023; CESP:T04F3.1
A:Experimental source: clone T04F3
C:Genetics:
A:Gene: CESP:T04F3.1
A:Map position: 5
A:Introns: 338/3; 417/3; 1957/2; 2358/3; 2400/1; 2529/2; 2585/1; 2700/3; 2744/1; 2877/3

Query Match 3.8%; Score 331.5; DB 2; Length 3147;
Best Local Similarity 18.8%; Pred. No. 7.7e-06;
Matches 349; Conservative 283; Mismatches 666; Indels 561; Gaps 82;

OY 2 NSTPS--KLPTDKISHILOPOSSASITFNSPTKPLNPRNRSKPSLDPN----- 50
Db 422 STLPFRNRRAPD---LELE-----DLEN--PKFSPHSANSKPPRRNRRHPPSAS 469

```

QY	51	-----	SSSDY	57
Db	470	VAHSSSFVDESPOQHVTYTTIDRNOYTPVTTTTTKMRESGPLLHNSHOLLSSDNW	1111:	529
QY	58	SE-----ODQEKKEKKDTAFOTSPD--BNFDLDSIDIOQTIOHQOQPOOQOOLQOT	1111:	110
Db	530	AOMNMDASEMENNNSKRKSSILSTSSAKKASARRISVDELTRPEKRRQAVPLPS	1111:	589
QY	111	DNMLIDESOTPM-----TSLDLTKON--PYDKNV-ENHAPTYINTSPKNSIMK	1111:	160
Db	590	DVSDIDLTALTLPKKEEVQSEKRFVITRRORADVADIDFEKDRPTI-----	1111:	638
QY	161	ATPKAPKVALFYTUNBEIHHPYDNRVEEDQSOQKEDSVEPLIOHQKDESOQFVNS--	1111:	218
Db	639	APPSCKRKSTATNOQETILEDVEKHEIDSKVSTJTN---LNDESMETRNNDSD	1111:	695
QY	219	-DEDTN-----ASVPPRPLHTTKPT---AOLLKNNENVSEPALDMLKRE-	1111:	264
Db	696	SFDEVDNPRNORLVEIPFSEPPRTSTATIOLESSDVAGENSE-NKRP--VISMRSKEI	1111:	752
QY	265	-----NFSULSDEKVNLYLSPTNNNNSKNVSDMSHLQNO	1111:	301
Db	753	AKKEKDAORSGFVIIPHSKEIIDESVISMDFN-----TPHDCRFPDIDA-----	1111:	801
QY	302	DASKNKTE-----NIHNSFLAKARKNDIENPLSLTNADISL-----	1111:	340
Db	802	DSKHTSDSDREVSTYINLDNV-FPREEPK-----LVAKDCEIEAEERIGRKIOF	1111:	854
QY	341	-RSSGSSQSSLOSLRNNRYLES-----VPSGPKVNGSLNIOIKF	1111:	383
Db	855	ERTIGEQISNSSEPTEDEMDEKDHRTSAVSIDLKVFQGTAKPE-----ND--EF	1111:	906
QY	384	SDEVESLPLKDLSDRDLLETTKEHDA--PENNNENFIDAKST-----NTNKQGLVSSD	1111:	436
Db	907	DEKIRGIAEPRKQKEVORSGVATSHSGKHIFEDSISMDFVNTSQ--KYKSD	1111:	963
QY	437	HLDSFDYSYN-HTQOSILNLNLSAQOISLMLKOROTOBEOQOAAPEPETSFSN	1111:	495
Db	964	KLSPERPERVEVSTATMNLDNISASGIATR--EENFNVALEEERIOKRPFEKTTN	1111:	102
QY	496	IKVQOEPSNLEFY-KVTIKKEPVSATIEIKAPREFSSRLIRKND-----IAPADI	1111:	549
Db	1022	LEIQEYVLTRKEVDNSDVAEHRSANVIDLEKVFIIHSSKKRKNDEKIRGIAFEPT	1111:	108
QY	550	HPKKEANEANSVDEPTOALLKALNDBESOTTONSTRKMSIRFIHDSMDKLEDSNDCRED	1111:	609
Db	1082	KOEKEAQSIVIE-----TSOSNS--RIFEESISMDFVFNLSHN	1111:	1120
QY	610	NDDISRFKSDILNDVQSOTSDIIOCKYGNSSSETITRTLAAPRSDNNKDKNSKL-----	1111:	664
Db	1121	ESQVSEITEADPDLDVLITSTFTINVAIEKIIDDYKT-----DSNVEEKEQVALRD	1111:	117
QY	665	-EDPANNESSLOQLLEVPHTRKEDDSILANSSNIAPPELTLPVVEANDYSFENDYKTFD	1111:	722
Db	1175	EFKRPTPEONLOKQEFELTKKEEYSV-----KMERNTSAVSIDLKQVD	1111:	1216
QY	723	AVSPEESLSEHETDKPILNFTSIHMKOEKKOHIIHKVPQOILASVQYKKNQDESV	1111:	782
Db	1219	-QSSKETTVS--NETDEKIRGIAEPRKQKEVORSGVAF--TSHSGKHIEDSNI	1111:	127
QY	783	TSDDYKATPNAIQKFKFEVNVMSRRVYSPMDLNVSQLPELSEDSGKDLNFANYSN	1111:	842
Db	1272	SMDOVF--NTSQKXKSDKELSSPERVTEPEVSNATNM--LDNIIFASG-----IATREEN	1111:	1322
QY	843	TN-----PRKSTPLSTKNVLSINDNPNVVEPEPKSYAETIRNARLSANKAP	1111:	892
Db	1323	TDVLEEERLOKRYEEF-----KKTENLEIOQEVVLTKEEVDNSGVKCHRTSAVINLD	1111:	137
QY	893	NOAPRPLPQORPSSTRSNKNSKRVSRFPVFFELRTS-----SALAPCDMN-----	1111:	939
Db	1378	DVF-----IORSKHPENDEDEKIRGIAEERTKQKEAORASVVIETSQSKHIEDK	1111:	1433

```

0Y 940 -----DIPDDGAGKPTIKAGGMKT-LPSMDKDDYV--RLNKKKGTOQDEY---INA 987
Db 1432 SNISMDVEYNESONGKODSSNIDMKETDMDPEKERDQRYVDHNRDKKPRPNCEFEPTENG 1491
QY 988 KLVDPKPKNSIVT-----DP---ED-----RYEELQOTASIHNAI 1020
Db 1492 SKISNEPKOISITITNLDNVPFTEEPKVLVADNCEIEAEFEERIKRKIKOFERTTGOEITL 1551
QY 1021 IDSSIYGRDSTIDMILPYLSDELKRPPTALLS--ADRLFMEOEYHPLRSNSVLYHPRGAG 1078
Db 1552 KNS---EPRADEPT-----SDE-KKHRTAAVSIDLDKVFYO----- 1582
QY 1079 AATNSSMLPEPPELINSIPARVNS-----NSDNLVASICNASTISFNOLDKNFDD 1128
Db 1583 ----GTAKKREND-EPDEKIKRGIAEFERSKOEYQORSVAGVAFETSHSIIKIPDESINISMD 1638
QY 1129 QATIGOKIOEOPAKSANSIVYRGDDGLASABETPRTPPKKESISAKKLASASPPKSP1 1188
Db 1639 VENTSOYKK-----SDEKLSPEKTYEP-----EVSATINLMDNI 1673
QY 1189 KIGSPRYVATKKKNSIAGIEP-IPKATHKPKKSPFOGNEISNHKV--RDGSISSGSEHOQ 1245
Db 1674 IRASGATATEKNKDYLEEERIRQKRVEEFKTTTEMLYLOKEVULLKEEG-----DNSDVAD 1729
QY 1246 HNPMSVY--PSQYDATSVPRDENKDVQHKPR-----EKOKOHNNHNNNNHNNKOTDI 1298
Db 1730 HKASAANVIDLDVFIORSSKHRENDDEDEKIRGIAEFERTKOE-----KEAORSTV 1781
QY 1299 PCGVDEIDPVGLOERKLEFRVLGINKININLPDINTHKGREFLTLDNGHCVATPRYND 1358
Db 1782 -----IEFOYSSKDMF-----NESDISLDVVPNTSOXKDSPEKLSSPRYVE 1823
QY 1359 -----DHNVAIKKEFELTVADSL-----FITLKASYEKPR-GTLVEYTEKK 1400
Db 1824 PEVSTATVINDNNVALSKERKKENNETOEEOEOIOKRVEEFESEDEOKLOKSIELTKEE 1883
QY 1401 VKSRNRILSLBESKODIITTKFVTEYKOTWANKFAFDGSAFCYIDLOQFEDQITGKA 1460
Db 1884 CTSDEKELKTYSGS-----IDDKVFIQSSSK 1911
QY 1461 SOFDLNCFMETMNSGNOQPKRGKPYKIAOLEVMKLVPR---SDPREILPTISRSAY 1516
Db 1912 PRND-----ESDERINRGIAEFERTKOEKRAQRSVVEFSPSKHISDESISMDIELFSRS 1967
QY 1517 ESINELANNEONNVPFEGVLHOGEGDCPIFKKRFKLMGSLASHSELSHKTRAKINLSKY 1575
Db 1968 Q-----DNKSTSMFE-----KSGSIP1-----IYLBGEKEVASASININGV 2004

```

RESULT 9

B89921
 Hypothetical protein ebha [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: B89921
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
 ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.: Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsutsu, K.
 Lancel 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: B89921
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-6713 <KUR>
 A:Cross-references: GB:BA000018; PID:g13701232; PIDN:BA042527.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: ebha

Query Match 3.8%; Score 329; DB 2; Length 6713;
Best Local Similarity 18.2%; Pred. No. 2.7e-05;
Matches 326; Conservative 303; Mismatches 661;
Indels 502; Gaps 81

QY	24	SSAJSLENSPTKRLPLFPRTNKPSPSLDRNSSDPYTSSEDOKEKREKKD-----	71
Db	2695	TTAADVTAATQVRV-----NMTSLNGDTJNLAT-----AKQQAALRQMHLSDAQ	2741
QY	72	-----TAFQTSFERNFPLDINSID-IQOTIQHOQOOPQOQOOL-SQTDNLI	115
Db	2742	KOSITGGQIDSAQTQVTGVSQKDNATNLIDNANMOLRMSINKBEVAAQSPYDADTDK---	2798
QY	116	DEFSQTEPMST---LDLTQRQNPYDKVNEBHNAPTYINTSPKKSIMKATPKASPKYAF	172
Db	2799	-QNAVNTAVTSAENIINATSQ-PLDRPSATQANQVNT--AKTLANGQANLAN-KQGET	2853
QY	173	TYTNPETHHYEDKNVEED-----OSQOKEDSVEPPLIQHMKD--	211
Db	2854	TANINRSLSHL--NNAQOKDLQNTQVTAAPNISVTNQYKTKAEQIDQAMERLINGIDQCKD	2911
QY	212	PSQFVNSPEDNANASVPRTPLHTTKPPEAO-LLNKNEVENSEPE-----ALDPMKLKRENF	266
Db	2912	KOSVNETDAD-----PEKQTAIYNNAVTAAENITIQANNGTNAQOSOEALSLVTTTKQ--	2964
QY	267	SNLSDLEKVNLYLSPNTNNNSKNVSDMSHQLIQDASKNKTNENI--HNLSPALAKPK	323
Db	2965	-ALNDKRV-----TDANKNANQTLSTLD--NLNNAQGAVTGNGINQAHYAEVTOAQ	3015
QY	324	NDIE-----NPLNSTNADISLRSSGSSQSLQSLRNDNRVLESVPSGPKVNPGLSN	377
Db	3016	TAEQELNTAMGLKSLNDKDTLTGSONFADA-----DPEKKN--AYN	3055
QY	378	DKIGKFSPEVEESLPLPDLSDKLETTKEHDAPEHHNENFIDAKSNTNKGQLLYSSDH	437
Db	3056	EAVNR-ANINILNKSTGTNRVPEQVDEA-----MNOYNTIKAAL--NGTON	3097
QY	438	LDSPDRSYNHTQOSTILNLSASOS-----QISLNLKQROTQEOEQTAAPPEEETSF	492
Db	3098	LEKAOQHANTADIGSHLTNAQKALKQLVQOSTYVAEAGNQNQKANNVDAAMDRLQSI	3157
QY	493	SDNIKVKOEPKSNLEFYKVYITKKEPVSATETKAPKKEFSRILIKN-EDETAEPADJHP	551
Db	3158	ADNATTKQ-----NONYTDASPNK-----KAYANNAVTTAGGIDQJTTNPS-LDP	3201
QY	552	KKENANSHVEDTALLKKALNDEESDTQNSTKMSIRPHIDSMKLEDSNDGQRED-N	610
Db	3202	TVINQAGQVSTS-----KNALNGENLEAAKQOATQSL-----GSLDNLNNAQQAHT	3250
QY	611	DISRFEKSDILINDVQSOTSDIGKAYGNSSEITTTTL-----APPSDNDK	658
Db	3251	NOINQAHYVDEANOIKQNAQNLNTAMQNLKQALDADATATATYNTPTDQAQQAQVNTAV	3310
QY	659	ENKSGLEDPAN-----NESIQOOLEVPHTKEDSDILANSS---NIA	696
Db	3311	TAAENITISKANGNATQTEVEBOAIQOVANAKQALNCANVQAHKADDEATLAINNSDNLQA	3370
QY	697	PEPELTLEPVANDYSSFNQYTKTF-----DAYSSFEELSREHETDSKINFISTMHQOE	752
Db	3371	QKDALKQOVQNAATTVAGVANNKYQTAQELINNAAPOLQKGLADKQJTKADG-NEVNA--DSD	3427
QY	753	KQKKQHJHVPKTOJIASYOOKNEQESRV--TSDKVKJPNALQFKFKEVWMSRRVVS	810
Db	3428	KQN-----AYNOAVAKAALISGTPIDVYVTPSEL-----TAALNKVT	3466
QY	811	PDMDLNVSOELPELSBEDSGFKDLPNFANYSNNTNRPSFTPLST--KNVLSINDIDPNV	868
Db	3465	QAKNDLNL-----GNTN-----LATQAQNVQHAIDQLPMLN	3496
QY	869	EPPEPKSTAETIRNARLSANKAAPNQAPLPPORQPSSTRNSKNKVSFERYPTEIRRT	928
Db	3495	QAOQREYSKQITQOA-----TLVPYNVAIQOQAATLTLDAMTQLOKGIANKQKQIGS	3544
QY	929	SSALAPCOMYNDI-FDDEGASRFTIABQM---KTLPSDKXDQVYRILNNAKGVTQOEY	984
Db	3545	EN-----YHDADTKQATADNAVTAABEELIKQTNPTMDPNTIQOALTKYNDQNALN	3597

Db 1947 TNDDEKITSKINDDFNITLIDKQ_KGKLLNNPNVDLNRKIKNE-----TKILEKOKSHMS 1000

QY 1061 QEYHPL---RSNSVLVHFGAGATATSSMLPE--PDFELINSBPAVNS-----NS 1105

Db 1001 KIONNMKKMKKNTFPNNNGISSTSSISSSTKNNFKDCGIEK-NKMLSLMKYTCIRKNMNV 1059

QY 1106 DNVAISGASASTISFNQILMNFDDATIGKIOEOPASKSANTVRGDDGLASAPETPRTP 1165

Db 1060 DSIKINDKA-----DLYDKKTSFNDINR--AARGMNFKKHD-----VP 1097

QY 1166 TKKESISSKPAKLSSASPRKSPIKIGSPVRYLK-----KNGSIAGIEPIPK 1211

Db 1098 NKNMNVDTNKKG-----RVFNPTLNNYRNRYINSKNNKYNKGMVGIKKIYVL 1146

QY 1212 ATHPKKSFQGNELSNKVRDGGISPSGSEHQHNPMSVSPQYDTATSTVDENKDV 1271

Db 1147 LKEKKK-----SLHPEGEVADKKILNSYNDKYLIEKDGFKDI 1183

QY 1272 QHKPREKOKHHHHHHHHHHKOKTDIPGVVDELIDVGLDERGKLFPRVLGIRKI----- 1327

Db 1184 INEEMEKTK-----NNKKRYIK--SNSIPPI-----IKKIERKS 1216

QY 1328 MLPDFINTKGRFTLLDNGVH-----CV-----TTPPEYMMDDHNVAGKEPEL----- 1370

Db 1217 NDNDNDNDKNNNDNINSNNNNNSNNDKCLFLEKEDRVALKNNNIIVNTTMMFRKQNSCD 1276

QY 1371 TVADSLFELITLKASYEKPRGTLVEVTEKKVYKSNRLSRLGSKDIT----- 1418

Db 1277 NNTTSLAKKKMLINSSEKANSNSQNNNEKKEMSYEFEMLANEKKKEITIEKKDEEKKGEI 1336

QY 1419 -----TTTKPVPEVKDTMANKFAPDGSFACYIDLOOFEOQITGRKASQPDNCFNME 1472

Db 1337 EKSEHANDSKMKDELDLDAKPTMLQPLSSF-----NLKQNEKKY--EQSDFLVDKY----- 1385

QY 1473 TMSNGNSQPM-----KRGKPKYK-----AOLEVKMLVPRSPREILPTSIKSAVESINE 1521

Db 1386 PIGNRGTGLVFKRAIIKKKEENKVKALKVMAKDTIMSLNIEQVLEKIT---IQASLKHINI 1442

QY 1522 LNEONNYFEGLYHOGGDCPLFKRRFFKLM----- 1552

Db 1443 L-ELIAFEED-----KTRLFLLELALANGSVRRMKMKQKQPLNEEVALYVF 1488

QY 1553 -----GTSLAHSEISHKTRAKTINLSKVVDLIYVDKENIDRSNHRNFSOVLILDHNAFKIKF 1608

Db 1489 QIADALSTLVHFNFIHTRDLKPDNI-----LIHYSMEHLNKKIYK-YGVIKLADFEFSCQL 1542

QY 1609 ANGE-----LIDFCAP 1619

Db 1543 KKKRKRSTFCGTTIDMPP 1561

RESULT 11

JC6009

surface-located membrane protein Imp3 precursor - Mycoplasma hominis

C:Species: Mycoplasma hominis

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: JC6009

R:Label: J. A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.

J: Bacteriol. 178, 2775-2784, 1996

A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis imp gene system

A:Reference number: JC6009; MUID:96213016; PMID:8631664

A:Accession: JC6009

A:Molecule type: DNA

A:Residues: 1-1302 <LAD>

A:Cross-references: EMBL:X95601; NID:g1197335; PIDN:CA64858.1; PID:g1197336

C:Genetics:

A:Gene: Imp3

A:Genetic code: SGC3

C:Superfamily: surface-located membrane protein Imp3; tetratricopeptide repeat homology

C:Keywords: duplication; membrane protein

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-1302/Product: surface-located membrane protein Imp3 #status predicted <MAT>

F:957-992/Domain: tetratricopeptide repeat homology <TPI>

Query Match	3.8%	Score 324	DB 1	Length 1302
Best Local Similarity	20.6%	Pred. No. 4,9e-06		
Matches	281	Conservative	221	Mismatches 463; Indels 399; Gaps 70
QY	9	LPIDK-----	HSILOLOPOSSASIFN-----	SPKPLN 37
Db	150	LPIDKRTTNLENLAKKELLINKINAERELQ-----	SKIFNEKQELKRVLDLEDTKEVD	204
QY	38	FPRTRSKSLDPNSSDPTVTSFQODEKGEKKEDTAFQTSFPRRNDLNSIDIQOTIQHQ	97	
Db	205	FTK-EQKVFLETN-INTSSIEDINKIIIEKATSSILTS-----	KILNTRK	248
QY	98	QOOPOQO-----QOOLQOTNNLLIDERSFQTPMTSTLDLTQONPLVDKVNENHAPTYINTSP	153	
Db	249	QOELQEFENIKKLDODFINTKLNDAKYQSIKKALD--KIN-SLNGINKNSITKEIKAQO	305	
QY	154	NKSIKKATPKRASPCK-----	VAFTVTPN-----	EIHHPDNEY 187
Db	306	NALI--KAKEEGEKEKELDQONIKDTLKETINNAKEPKLLIDNDQKIVDLKSLNDNEI	363	
QY	188	EEDDSOQKE-DSVPEPLIQHOMKRPDSQPNYSDEDTNANVPPTPLHTTKPFFAQILANK	246	
Db	364	SKAEQSLSKDKESME-----	SANDL-----	LNTKLEKELINK- 397
QY	247	NEVNSEPPA-LTDMKLRKRENFSLDEKVN-----	LYLSPTNN-----	NNSKNVSD 292
Db	398	--FNOEKRAKFFELQOTKRNINENPLTDEKKNPNATLVKDLTNAKDAKKSVTNNSKSD	455	
QY	293	M-----DSHLQNLQDASKNNKTNNENIHLSPALAKPKNDIENPLNSL-----	333	
Db	456	IIAANEALIQALADA--NKAKQVDEANKSIEQNALALIDKANTLLPQLNDNDSEIVKAK	513	
QY	334	--TNDISLRSSGSSQSSLSLRNDRVLESVPGEFKKYNPGLSLNDGKIGFSDEVESL	391	
Db	514	ESLNEIETNNANKAVNO-----	NDNASQSAKS-----	SLDDKVTYKIQIONQLE-- 555
QY	392	LPRLDSRLKLETTKREHDAPEHNENPI--DAKSTNTKQOL--	IYSSDDHLSFDKSYNH	447
Db	556	--FNKKDKAEKLEQTRKIDINFLTDVAKNNPVATLVKDLTNAKDKKSVYKSSNK	611	
QY	448	TEOSILTNLSASQOISLNALAEKORQDEQOTQAABEEETSFSDNKTIVKQEPKSNLE	507	
Db	612	SE--II-----	ANAEKQ-QLADKAKVAKQDID-----	EANKSIEQSLDSTT 651
QY	508	FVKVTIKKEPVASATEIKAPKREFSSRIILIKNEDETAEPAIDHPKKEENANSHEVDAL	567	
Db	652	NANQLLNKLVDSDKDIQAKTELISOEIGASQELNMLNPTSMQAKES-----	LDAK	703
QY	568	LKKALNDEESTPTQNSRTKMSIRPHIDSWKLEDSNDGREDNDOLSPREKSDILNDVQO	627	
Db	704	VTEI-----	TKLETENKDKVYKKELE- EKKRKDIDEEINTKTN-----	742
QY	628	TSIDIGDKYGNSSSEITTTLPAPPSDNDKSNKSLSD-PAANNLSLQOOLEVPTKED	685	
Db	743	-----	PKVSTILSELTSK-----	KDSKSVYNNSSKSDILETANTE-LKQALAKANTDK- 789
QY	686	DSILANSSNIA--PPEELTLPLVEANDVSSPND-VTKTPDAYSSPEBSLSRBEHENDSKIPI	742	
Db	790	---AQADNLAKSTKEQLN-----	NSISSANTLLAKLTDKDNNTIQQA-----	827
QY	743	NFISIMHQQEQKKHO-IHKVPTKQILASVQOYKKNQOESRVPKDKYKINMALOFKKREY	801	
Db	828	-----	TELEKEVQAKAQAAVANNATJASQAKSSLDKAVETLTKKLET--	FNDKQV 877
QY	802	NVMSRRVSPDDMD-LVNSQFLPE-----	LSDESGFIDL--	NFANSYNNTPRPSFPLTKED 854
Db	878	KFKLEJQTRKQDEIINTKNTPOVSTYLSIELTSKRSKSNSTNNSKSDILETANTELQO	937	
QY	855	KNVLSININD--PNVVEPEPKSVAEIRNARRLSANKAAPNOAPPLPQROPSSTRSNGN	912	

Db	938	ALAKANTDKQADNLNARSTKEQLNKSISANTLLAKLT-----DKONTIOQAK	985
Qy	913	KRVSPFRVPTEFIRKSSALPCDMYNDIFDDGACSKPPIK-AEGKTLPSMDKD-DVK	970
Db	986	TELEK-----EVOKANOVAAS---NNTASMOASASLDAKTEITTKLLETENKOKDVK	1035
Qy	971	--RIINAKGVIOQDEXIN-----AKLVDCPKKSNIV-----DPEDERYELO	1011
Db	1036	FRELEQTRKDI--DEFINNTKNPNYSTLSELTSKRDSKNSITNSNKSDIETANTELK	1093
Qy	1012	QTASHTNATIDSSITGRPDSISTDMLPYLSEDEIKKEPTAL-----LSADRLPMEQE	1062
Db	1094	Q--ALAKANTDKA---QADNLNARSTKEQLNKSISANTLLAKLTDKDKNTIOQAKTELEKE	1148
Qy	1063	V---HPRLSNGLVHPGGAATNSML-----PEEDF-----ELIN	1095
Db	1149	VOKANOVAASNNTVSMQAKSSELDTKVTEITTKLETFNNKDEKKEFELKKTROIOIEFIN	1208
Qy	1096	SPARVNSNNSDNA-----ISGNASTISFNOLDMNFDDOATIGORIOEQPASKSANTVR	1149
Db	1209	T-NKNPNPSELISOLTSKRDSKNSVTNSNKSNDIE---TANTELKQ--ALAKANTDK	1260
Qy	1150	GDDDLASAPETPRPTTKE---SISKRPAKUSSAPKRSPIK	1189
Db	1261	AOADNLNARS-----TKEQLNKSISANTLLAKLTDKDKNTIOQ	1296

Db 1780 YNNLKKYCDMEHEDSLMLIKHLISVLKDEKENIKNNEGSEIDNLLID-----KNMKS 1835
 QY 1405 RNRSLRLEFGSKDITTTKFEVPTVDKMTANKFAPDGSFARCYIDLOQFEDQITGKASOPD 1464
 Db 1836 DQENE--NEKHDLNNIEYI-ENLSEHLRHKY-----DIEMIPIDIVGNIRLTK 1881
 QY 1465 LNCFNEMET-----MSNGQPH-----KRGKPK-IAOLEVKNLY-----1499
 Db 1882 LFYINNIKNNYSYLKNNFPPKYKGRKAKGMLIMIKRKHEDNLMKKNGNEKGT 1941
 QY 1500 -----PRSDREILPTIRSAYESINELNNEQNNYFEGYHDEG-----1539
 Db 1942 NNNKKNNKEDNDYINNN-----NNNNNNNNNNINNGYKRMQNNINPNLYETIKNN 1996
 QY 1540 -----DQPIFKRFFKLMG-----TSLANSEI-- 1562
 Db 1997 FNLHYRWSLFKYCPVSYINDNKIEGKKFCVYIKNKIYMLNNDYIEFLKHPFYA 2056
 Db 1563 SHKTRAKILSLVY-----DLIYV-----KE 1584
 Db 2057 SKRVESVCSFLVYFPNMDIOKTIILYLETYEKLHIDIGYFKQYFPLSHQDKRQHN 2116
 QY 1585 NIDRSHNRNFSVLLD-HAFKI-KFANGELID-----FCAPNK--HEMKIWNLOE 1633
 Db 2117 NKKSNNDODVONKIEDAIFIEIHRTEGKTISPFIYVYICIMHKKTQFMNIT--IKE 2174
 QY 1634 IYRNRFRQRPVNLMLQOOQ 1654
 Db 2175 II-KSIILKLNHFNLLIMKK 2194

RESULT 14
 ORMSPL
 microtubule-associated protein MAP1B - mouse
 N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei
 C:Species: Mus musculus (house mouse)
 C:Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text change 01-Sep-2000
 C:Accession: S07549; S44387; A33645
 R:Noble, M.; Lewis, S.A.; Cowan, N.J.
 J:Cell Biol. 109, 3367-3376, 1989
 A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contains
 A:Reference number: A33645; MUID:90094539; PMID:2480963
 A:Accession: S07549
 A:Molecule type: mRNA
 A:Residues: 1-2464 <NO>
 A:Cross-references: EMBL:X51396; MID:952999; PIDN:CA35761.1; PID:953000
 A:Author: Manchez, C.; Padilla, R.; Paciucci, R.; Zabalza, J.C.; Avila, J.
 A:Ch. Biochem. Biophys. 310, 428-432, 1994
 A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.
 A:Reference number: S44387; MUID:94234720; PMID:8179328
 A:Accession: S44387
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 653-663, 'IC' <SAN>
 C:Superfamily: microtubule-associated protein MAP1B
 C:Keywords: microtubule binding; phosphoprotein; tandem repeat
 F:589-786/Domain: microtubule binding #status experimental <MTB>
 F:589-592, 639-642, 649-652, 655-658, 660-663, 668-671, 674-677, 679-682, 683-686, 687-690, 691-69
 R-K-E/-D-X)
 F:1861-2064/Region: 17-residue repeats
 F:91, 116, 351, 868, 1124, 1153, 1168, 1208, 1662, 1877, 1918, 2003, 2030, 2054, 2083/Binding site: ph
 F:147, 969, 1336, 1562, 1563, 1702, 1708, 1990, 2057, 2063, 2419/Binding site: phosphate (Thr) (cc
 F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 3.6%; Score 314; DB 1: Length 2464;
 Best Local Similarity 17.7%; Pred. No. 3e-05;
 Matches 341; Conservative 319; Mismatches 736; Indels 534; Gaps 92;

QY 2 NSTPSKLP-IDKSHLQ--QPOSSASIFNS-PTKPLN---FPRNKSPLDPNSSS 53
 Db 492 NSTQVNLGEGLEKHLFLKPLATOKDLTGQVTPPVKQVLEKLRADSRSLKP--AT 549
 QY 54 DTYTSDQDQEKKEKKATPOTSPDRNFDLNSIDIOQTIOHQOQFQOQOQLSQTDMN 113

Db 550 KPVASKSVKKEKEETPEVTKTSQVEKT-----PKVESKEKVLVKDKDPVKTSKSPVTEK 605
 QY 114 LIDEFSPMTSTL--DLTKNP--TVDKYNNENAPVYI-----NTSPNKSIMK--AT 162
 Db 606 EYSSKEQSPVAEVAEKQATSKPVTKQVYKIKLEKLEKKEEKKVKKEDKT 665
 QY 163 PRASPKVAFVTVNPEIHNYPDNRVEEDOSO-QKEDSVPEPL-----10H 207
 Db 666 PLKDEKPRKEVKKEL-----KKEIKKEERKELKVEKKEPLDKAKKEVKKKEKVK 721
 QY 208 QMKDP--SQNTYDEDTNASTVPTPLHTTKPTF--AQLLKNNEVNSEPALTDK-- 260
 Db 722 EKEPKKEIKISKDIKS--TPQSDTKPKPALPKVAKKEESTKKEPLAAGKLKDKG 778
 QY 261 ---LKRNFNSN-----LSDEKYNLYLSPNNNSKN 289
 Db 779 KVYIKKEGTTAAATAVGTAATTAAYVAAGIAASGPVKELEAERSLMSFEDLTKEE 838
 QY 290 -----VSDMDSLQNLQADASKNKTNE--NIHLSFALKAPNDIENPLNSLTNADI 338
 Db 839 ELKAEIDVAKDIKPOLLEIDEEKLETKQPEAVYIQETEVSGSASPDGITTTE- 897
 QY 339 SURSSSSSSLSQSLRDNKRVLESYPGSKVNP-----GLSNDGIKGSDE---VVESL 391
 Db 898 ---GEQEQO-----TPELEPVEKQGV---DDIEKFEDEGAGFEES 934
 QY 392 LPRLSDRLKLETTKEHADPEHNNENFIDAKSTNTKQGLLVSSDDHLDFDSYNNHTOS 451
 Db 935 ETGDI-EEKAE-TEEAPEEEDGE---DNASSASKHSTEDDESKAADVHLKEKRS 989
 QY 452 IINLNSASQISLNALEKOROTQEOQOAEPEEETSFSDNIKVKQEP-KSNLE--F 508
 Db 990 VYSGCDRAEEDMD--DVLEKGEAQSEE-----EEEBEKADAREBEVPEKTEADVY 1042
 QY 509 VYVITKKEVSAATE-----IKAKREFSSILIKKNEDETAEPADLHPKE 554
 Db 1043 MAVADKAAGVTEBOGYGLTSAKOPGIGQSPSREPASSI-----HDE---TLPGSE 1092
 QY 555 NEANSHVEDTDALLKALNDDESDTTONS-----TKMSIFHIDSDK 598
 Db 1093 SATATSDEN-----REDPEEFATISGTQSTIETSSPTMDESTROVMSDET 1144
 QY 599 LEDSDGREDNDISREKS-----DILNDVQSOTDIIGDKYGNSSSETT 644
 Db 1145 NNEETESQSEFVNITTKYESSLYSQEYKPAVASFNGLSGSKTDATQKOVNASAS-- 1201
 QY 645 TKTLPAPRSDNNDKNSKSLERPANNESIQ-----QOLEVPHTKEDDSILANSNAP--- 697
 Db 1202 --TISPSSMEEDKFSKSLRDVAYSEEEKELKASAEIDIKVDSDERLSPAKPSLSPSP 1259
 QY 698 -PEELTLPVVEANDYSF--NDVTKTFDAYS-----SFEESLSREH--ETDSKPINFISI 747
 Db 1260 SPIETK-PLGERSVNFSLPTEINIKVASBEASVSPGYQVAVEHCASPEKTELEVVS- 1317
 QY 748 WHKQEKKKHQTINHYPTKQIIAS-----YQQYKNQOES---RYTSDKVILNADIQPK 797
 Db 1318 -----PQSVTSAGHTPYQSPTEDEKSHLPTYESNAQAVPVSEFSE 1362
 QY 798 FKEVNVMSKRVVSPMD-----LNVQFLPELSDSQFKDLNFANYSNNNTR 845
 Db 1363 AKDEN--EKASLSP-MDEVVPDSESPVKVLSPLRSPPLGSESPYED--FLSADSKVIG 1417
 QY 846 PRSFTPLSTKVNLSINDNDPNVPEPEPKSVAEINARLSANKAQAPOAPLP----- 899
 Db 1418 RSESEPFEGKCKQGF-----PDRESVPYSDLTSTGLYODKQKEESTGTGHPKPEDG 1468
 QY 900 PQROPSTRSNSNKRVSRRVPTFEIRRTSALACDQVNDIIFDDFGAGSKPTIAEGMK 959
 Db 1469 PEKRTSDVETMSQAL-----ALDERKLGDSVSPQTLDVDSQFGSF----- 1509
 QY 960 TLPSMDKDDVRIKLAAGGVQDEIYINAKLVQDKKKNSIYTPDPRDELELOQTASTINA 1019

Db 1510 -----KEDTK--MSISEGTVSDK--SATPVDEGV-----AEDTSHMEGVASVSTA 1551
 QY 1020 TIDSSYGRPDSISTDMLPLLSDELKRP-----PTALLSADRPFMEQVHPLRSN- 1069
 Db 1552 SVATSSF--PEPTDVSPGLAEVSGPHSTEVDDLSVSVQPTPTTEEMSPSKSEEC 1609
 QY 1070 ---SVVHPGAGATNSSMP-----EPPDFLIN 1095
 Db 1610 PRMMSISPPDFSPKTKAKSRFPVQDHRSEQSSMSIEREQSEPSHMFANDFSRQSDHTTLC 1669
 QY 1096 SPARNVSNNSD-----NVAISGNASTI-----SFNOLDNMFDDQATIGOKIOBPA 1141
 Db 1670 ASYLHITENGPTEVDYSPCDIODSSLSHKIPTEEPSYTO-DNDLSLLSVSO-VEASP 1727
 QY 1142 SKANVVRGDDGLASAPETPRTPTKKESIS-----SK 1174
 Db 1728 TSSAHT-----PSQIASPLDIEDLSDVVPREMSLYASLASEKVOSLEGELSP 1776
 QY 1175 PAKLSASPRKSPKIKIGSP-----VRVKKNGS---IAGIEPIPKATHKP---KKSFGN 1223
 Db 1777 KSDISPLTPRESS-PLXSPFSDSTSAKETAHAHQSSSPIDATAETAEYGFRRSMFLD 1835
 QY 1224 EISNHNKVRDGGISPPS-GSEHQONPSMVSVPQYTDATSTVDENKDVQHKPREKQK 1282
 Db 1836 TMOHNLALNLDLTTSSVEKDSGKTPGDFNVAYOKPENAGSPDE-EDYDESQEKTIPT 1894
 QY 1283 HHHHHHHHKKOTDIPGVVD-----EIPDVG-----LQERGLTF 1319
 Db 1895 HDVVRYYEKTERT-IKSPCDSGSYETIEKTKTPEDGTYCEITEKTRPREGGYSY 1953
 QY 1320 RVLGKININLPDIN-----THKGRFTL-TLDNGVHCVTTPREYMDDHNAIGKEFELT- 1371
 Db 1954 EI-SEKTRKREVSQTYEKTERTSRLLDLSNGYDDEGHTLGCCSYETIEKITS 2012
 QY 1372 --VADSEETLLKASTEKPRGLVEYTEKRVKSNRLSLRSGSKD-IITTKFVPTV 1428
 Db 2013 FPESSESYSYETSTRTSPPTSAVCYETMEKIKTPQASTYSYETSRCYCTTEKKSSEA 2072
 QY 1429 KOTWANKFADGSPARCYIDLOQEQITKASQF-----DLNCNEMETMSNGQPKR- 1483
 Db 2073 KQ-----DVLCLVSSCEFKAPKTELSPFINPBLEWAGEEPPESEKPLQOS 2122
 QY 1484 -GKPYIAOLEVLMYVPRSDPR-----ELPSTIRSAVESINEINNEQNNFYEGYLH 1535
 Db 2123 GGAP-----PPSGGKQGGQCODEPTPTSVSEASAPOTDSD-----VP 2159
 QY 1536 QEGGDCPEFKRPFKLMGTSLAHSEISHTTRAK-INLSKVVDLIYDKENI---DRSNH 1591
 Db 2160 PETEECP-----SITADANIDSEDESETIPTDKTVTYKHMPPAPWQDSRPS 2207
 QY 1592 RNFSVLLLD 1601
 Db 2208 PRHPDVSMVD 2217
 RESULT 15
 T34418
 hypothetical protein f12f3.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34418
 R:Fulton, B.; Wohlmann, P.
 Submitted to the EMBL Data Library, July 1998
 A:Description: The sequence of C. elegans cosmid f12f3.
 A:Reference number: Z21521
 A:Accession: T34418
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3448 <FWU>
 A:Cross-references: EMBL:U080022; PDB:AAIC25865.1; GSPDB:GN00023; CESP:f12f3.3
 A:Experimental source: Strain Bristol N2; clone f12f3
 C:Genetics:
 A:Gene: CESP:f12f3.3

A:Map position: 5
 A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1
 Query Match 3.68; Score 313.5; DB 2; Length 3448;
 Best Local Similarity 18.44; Pred. No. 5e-05;
 Matches 344; Conservative 282; Mismatches 649; Indels 593; Gaps 80;
 QY 43 SKPSLDNNSDPTYSFQDQEKKEKKDPAFQSPDRNDPLDSIDIQOTIQOQOPO 102
 Db 330 SEESRDDKSYD-----EVDOSTYLEKKDDGDKSKPT-----KKLLIKKETPE 376
 QY 103 -----OQOQLSOTDNNLIDFESF-----OTPMST 127
 Db 377 SEQYTAAPRQOKISEVDQVSVAETEVGAKKKPRAEKPDTLSKAKKSKKSQSEPEAST 436
 QY 128 LD-LTKONPVTYKVENHA-----PYINTS--PN 154
 Db 437 EEKSTTEKPTNDKTSKSAEKKTVKPKREVTKGLEAKKPYEDKDKASOPSSKSSPT 496
 QY 155 KSIKKKAPRA--SPKKVAFVTNPEIHHYPDNVE-----EEDSQCKEDSVPE----- 202
 Db 497 DGKKKKOIPRALFTPDELISRFGDPTMHSETNITTTIRGREGSADAKTPLVEPLASVS 556
 QY 203 ----PLIOHMKDPSQFNY-----SDEDTNAYVPPTPLHTTKPTFAOLLKKNNEVNSE- 252
 Db 557 MKVFTLVESA-KEAKEFSFKRSETPDDKSKKKGLEP-----AKKSEKKDEPTAK 607
 QY 253 --PALTDMLKRENFNSLDEKVNLYLSTNNNSKNVS-----DMDSLQNLQD 302
 Db 608 QSTALELESKKKEVDESKISEQ-----PSDKNSFVGVKPAKAPETKKVSEIEE 660
 QY 303 ASKKKTENIHNLSFALKAPKNDIENPLNLTNADISLSSGSSOSS-LOSLRNDNYLE 361
 Db 661 VPKKKTIRK-----KTEKSD-----SISOKSNVLAADDKSKSD 696
 QY 362 SVSPGKRVNPGSLNDGKIGFSDVEVESLPRDLSRDKLETTTEHAPENHNENFTDAK 421
 Db 697 DVTGSKK-----TTEDQTKVATDSKLEK-----AADTKQIE-----TEVYVDK 737
 QY 422 STNTNKGOLVSSDDHLDSPRSYNNHTEOSLNLNLSASOSO--ISLNALEKQOTQOE 479
 Db 738 S---KKKVLKKTKEKSDSFISQKSETPPVETPKPASEKQKTAEVNAKKQKEVNDL 793
 QY 480 QTOAPEEETSFSFSDNFKVQOEPSNLEFVTKKEPEVSATEIKAPREBSYLIRKN 539
 Db 794 KREAEVAAKIA---DEKLKLEAAN-----IKK-----TAEVBAKKQ-KEKDEQIKL 838
 QY 540 EDEIAEPADIIHPKKENANSHVE---DTDALLK-KALNDESDPTTONSTKMSIRFHDS 595
 Db 839 ETEVVSKSAAEKLELEKQAKIKKAADAVAKKQKELNEKKLEAKKSAADKL----- 892
 QY 596 DWKLEDSNDGRENDDISREKSDILNDVSQTDIIGDKYGNSSSETTTTTLAPPSDN 655
 Db 893 --KLEE-----ESAASKRYSESVKFGEEKTKTAGETVQVSESEPTSK 935
 QY 656 NDKENSKSLDEPPANNESLQOQLEVPHTKEDDSIILANSSNIAPPELTLPVVEANDYSFN 715
 Db 936 TIDTKDVGATPRA-DETPKKIILIKKTEKSS----- 966
 QY 716 DVTKTFAYSSFEELSREHETDSKPINFISIMHKOEKQKHQIHKYPTKOILASVQYK 775
 Db 967 -----SISOKSATDSEKVS-----KQKQODE-----PTKAVSPTQVVT 1000
 QY 776 NEQES---RVNSDKVKIPNALQFKKFEVNVMSRRVYSPMDLDLNVSGFLELSDSGF- 831
 Db 1001 EADSKSKQKETDEKLDAELAAATKQEDAKSK-----LDAQKIRIKVSEDDAAR 1051
 QY 832 --KDLNFANYSNNTNRPSFPLSTKNVLSINDPNVVEPPEKSVAEINARLT----- 885
 Db 1052 KEKELN-----DKLKESEIATKASA---DKLKEAQAKAAEVAEAAKKQKOKD 1100
 QY 886 -----SANKAAPNQAAPLPQOR-----PSTRSNSKRVYSRFRVP 921

```

Db 1101 EQLKLDTEAASKKAAAEKLEKQAOIKKAAGADAVKQKQELDEKKNLEANKKSAAGKL- 1159
Qy 922 TFEIRRTSSALAPCDWYNDIFDDFGAGSKPTTIAEGMKTLPMDKDDVKKILNAKKGVQ 981
Db 1160 -----KIEESAASAKSQTVEEQAKLDAQTKAKTAKEQTKLEKDEKST 1201
Qy 982 DEYINAKLYDQPKKNSIYTDPEDRYEELQOTASIMATIDSSIYGRPDSIS----- 1033
Db 1202 KESESEKYDERKKKVKLKKTEKSDSSISQKSETSKTVESA--GPSESETKVADAAR 1259
Qy 1034 -----TDMLEPYLDELKPPALSLADRLMEQVHPLRSNVLVHPGAGATNSMLPEP 1089
Db 1260 KQKETDEKQKLEAEI-----TAKKSAD-----EKSKLEAESKLLKAAEVAANKQKEKDE 1309
Qy 1090 DEFLINSPA-----RNVSNNSDNVAISGNASTISFNOLD 1123
Db 1310 QLKLDTEAASKKAAAEKLEKQSHIKKAAEYDAVKKQKLEKQKRLSESAATKKADEK 1369
1124 MNFDDQ-----ATIG-----OKIQOPAKSANTYRGDDGLASA-----PETPRTPK-- 1167
Db 1370 LKLEEQKKAELALIEIQOEKLAQEOS--RLEDEAKKSAEKOKLESETKSKQTEEA 1426
Qy 1168 -KESISSKPAKL-----SSASPRKSPK-----IGSPRVITKKNGSITAGIEP 1208
Db 1427 PKESVDEKPKKVKLKKTEKSDSSISQKSAKSTYDAAEETLESDFNLVEKK-TVQKVEQ 1485
Qy 1209 IP-----KATHKPKKSPQNEISNH-----KYRDGSISSSGSEHQHNPMSVSPQYT 1258
Db 1486 SPDESTATIKRDPQAKTEIISQODGDDEKTTTDDCKPPKPEDESEATPKKRVYKKTKQKS 1545
Qy 1259 DAT-----STVPDENK---DYQHKPREKQKQKHHHHHHKQKTDIPGVVDEIPDYGLQ 1312
Db 1546 DSVASDASLADVSKLSDVDEKPKKVKL-----KKKTEKSDSVISETSSVD-- 1591
Qy 1313 ERGKLEFFRVIGIKINILPDINTHKGRTLLDNGVHCVTTPPEXNMDDHVAIGKEFELTV 1372
Db 1592 -----TIKPESEVIP-----TEKAQOMILHN-----RFS--- 1615
Qy 1373 ADSLEFILTAKASYEKPRGTLEVEYTEKV--VKSNNRLSLRFGSKDITTTKFPV---T 1426
Db 1616 -----TDSAVSESEPKNAHKDDTEKTTDDMTTRKSSAIFSDPOSISSTSEGRRR 1667
Qy 1427 EYKDTWANKFAPD-----GSFARCTYDLOQFEDQITGKASQFDLNCFNEMETMS----- 1475
Db 1668 RRRTGFAASKFASDTLALRGDNVEIEAELLAEDDTVTWKVNGKADADLNSRCHESHTEFRT 1727
1476 ---NGNQPMKRGKP-----YKIAOLEVKNL-YVPR-----S 1502
Db 1728 LIIDEVEPTDSGMEITATGCTESHITILKVEELPVDFVKYLPKRTSGKEQEVYISVTLN 1787
Qy 1503 DPREI-----LPTSIRSAVESINELNNEQNNYFEGYLHQEGGDCPIFKKRFPKLM- 1552
Db 1788 HPIDISKVYVWLKDKPRLINKDY-SIDTVGCSVSLTLRAKAYEDSGK-----YKVVV 1838
Qy 1553 -GTSLLAHSEISHKTRAKINLSKYVDLIYVDKENIDRSNHRNFSVDYLLLDHAKIKFA-- 1609
Db 1839 DGVDSTHLSIQKPYLK-NVSETKPVITVDDK-----QFSLLVAYDSNPEASFSMT 1890
Qy 1610 -NGELIDF 1616
Db 1891 VDGKDLEF 1898

```

Search completed: March 17, 2003, 12:26:55
Job time : 115.814 secs

FT	CAROHYD	638	638	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	670	670	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	691	691	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	817	817	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	838	838	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	841	841	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	1018	1018	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	1082	1082	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	1100	1100	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	1103	1103	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	1113	1113	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	1200	1200	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	1571	1571	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	CAROHYD	1593	1593	N-LINKED	(GLCNAC . . .)	(POTENTIAL) .
FT	DOMAIN	95	106	POLY-GLN	(GLCNAC . . .)	(POTENTIAL) .
FT	DOMAIN	283	286	POLY-ASN		
FT	DOMAIN	1283	1292	POLY-HIS		
FT	DOMAIN	1651	1660	POLY-GLN		
FT	SEQUENCE	1664 AA;	187859 MW;	BAZEF0DDC8196790 CRC64;		
Query Match						
Best Local Similarity		100.0%;	Score 8631;	DB 1;	Length 1664;	
Matches 1664;		Conservative	0;	Mismatches	0;	Gaps
OY	1	MNSTSKLLPLDKSHLLOLOPOSSASIFNSPTKPLNPRNTSKPSLDPNSSSDTYTSEQ	60			
Db	1	MNSTSKLLPLDKSHLLOLOPOSSASIFNSPTKPLNPRNTSKPSLDPNSSSDTYTSEQ	60			
OY	61	DOEKKEKKDPAFTSPFRNFDLNSIDIOTIOH000QPO000LOST0NNLIDENSEF	120			
Db	61	DOEKKEKKDPAFTSPFRNFDLNSIDIOTIOH000QPO000LOST0NNLIDENSEF	120			
OY	121	QTPMTSTLDLTQNPVTDKVNENHAPYYINTSPNKSIMKKATPKASPKKVAFTVNPETH	180			
Db	121	QTPMTSTLDLTQNPVTDKVNENHAPYYINTSPNKSIMKKATPKASPKKVAFTVNPETH	180			
OY	181	HYPDNRVEEDSOOKEDSVBEPPLLOHOMKDPQSNYSDEDTNNAVPPRLHTTKPTFA	240			
Db	181	HYPDNRVEEDSOOKEDSVBEPPLLOHOMKDPQSNYSDEDTNNAVPPRLHTTKPTFA	240			
OY	241	OLLNKNEVSEPALJTDMLKRENFNSLJDEKYNLJLSPNTNNNSKRVSDMDSHLNL	300			
Db	241	OLLNKNEVSEPALJTDMLKRENFNSLJDEKYNLJLSPNTNNNSKRVSDMDSHLNL	300			
OY	301	ODASKNKNTNENHNLSFALKAPKNDIENPLNSLTNADISLRSSGSSQSLSLRDNRYL	360			
Db	301	ODASKNKNTNENHNLSFALKAPKNDIENPLNSLTNADISLRSSGSSQSLSLRDNRYL	360			
OY	361	ESVPSPKKVNPGLSLNGICGFSDSEVYESLPRDLSBDKLETTKEHAPRHNENNFIDA	420			
Db	361	ESVPSPKKVNPGLSLNGICGFSDSEVYESLPRDLSBDKLETTKEHAPRHNENNFIDA	420			
OY	421	KSTNTNKQLLVSSDHLIDFSRYSNHTQSTLNLNLSASQISLMLAKORQTOEOO	480			
Db	421	KSTNTNKQLLVSSDHLIDFSRYSNHTQSTLNLNLSASQISLMLAKORQTOEOO	480			
OY	481	TOAAPEEETSPSONIKYKOEKPSLSEVKTATIKKEPSAETIKRKEPFSRILIRINE	540			
Db	481	TOAAPEEETSPSONIKYKOEKPSLSEVKTATIKKEPSAETIKRKEPFSRILIRINE	540			
OY	541	DEIAEPADIIHKKENEAHSVHEDTALLKALLNDEEEDTQNSTKMSIRFIIDSDMKLE	600			
Db	541	DEIAEPADIIHKKENEAHSVHEDTALLKALLNDEEEDTQNSTKMSIRFIIDSDMKLE	600			
OY	601	DSNODREDNDNDISFEKSDILNDVSOISDITIGKRYGMSSEITTKTAPPSDNDKEN	660			
Db	601	DSNODREDNDNDISFEKSDILNDVSOISDITIGKRYGMSSEITTKTAPPSDNDKEN	660			
OY	661	SKSLDEDPANNESLQOOLEVPPTKEEDSLIANSNTAPPEELTLPVVEANDVSSFNDDYKT	720			
Db	661	SKSLDEDPANNESLQOOLEVPPTKEEDSLIANSNTAPPEELTLPVVEANDVSSFNDDYKT	720			
OY	721	FDAYSSFEESLSREHETDTSKPIFNISIMHKOEKOKKHDIHVPTKQIIASVQOYNEDES	780			

[illegible]

RESULT	2
BUD4_YEAST	
ID	BUD4_YEAST
AC	P47136
DT	01-FEB-1996 (rel. 33, Created)
DT	01-OCT-1996 (rel. 34, Last sequence update)
DT	30-MAY-2000 (rel. 39, Last annotation update)
DE	Bud site selection protein BUD4.
DE	BUD4 OR YJR092W OR j1905.

OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9629555; PubMed=8707826;
 RA Sanders S.L., Herskowitz I.;
 RT "The BUD4 protein of yeast, required for axial budding, is localized
 RL to the mother/BUD neck in a cell cycle-dependent manner.";
 RL J. Cell Biol. 134:413-427(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ramezani Rad M., Kirchraeth L., Hollenberg C.P.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
 CC -1- FUNCTION: CO-ASSEMBLES WITH BUD3 AT BUD SITES. BUD4 AND BUD3 MAY
 CC COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)
 CC DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR
 CC ESTABLISHING THE AXIAL BUDDING PATTERN IN G1.
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N-TERMINAL
 CC SECTION DUE TO FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U41641; AAB17116.1;
 DR EMBL: 249592; CA89620.1; ALT_FRAME.
 DR EMBL: 249591; CA89619.1;
 DR SCD: S0003852; BDD4.
 DR InterPro: IPR001849; PH.
 DR Pfam: PF00169; PH.1.
 DR SMART: SM00233; PH.1.
 DR PROSITE: PS50003; PH_DOMAIN.1.
 KW ATP-binding.
 FT DOMAIN 181 189 ASP/GLU-RICH (ACIDIC).
 FT NP_BIND 1175 1182 ATP (POTENTIAL).
 FT DOMAIN 1302 1413 PH.
 FT CONFLICT 291 328 PTASVEDINISTNLPGADSSONNPVTADALINDVY ->
 FT ANSLGGFKFHESSRVLLPAKIIISLKRMLLTMTLC
 FT (IN REF. 2).
 FT CONFLICT 340 340 D -> E (IN REF. 2).
 FT SEQUENCE 1447 AA; 164485 MW; 1D056B9FF1B7067B CRC64;
 SQ
 Query Match 8.3%; Score 719; DB 1; Length 1447;
 Best local similarity 20.9%; Pred. No. 3.5e-21;
 Matches 357; Conservative 283; Mismatches 629; Indels 440; Gaps 70;
 OY 106 QLSQTDNMLIDFESFOTPMT-----STLDLKONPTVDKVEN--HAPTYINTS 152
 DB 18 EMEQSEKNTQNGSEDTPHNMKLPLOEIGDDTMEMLVHNHTRSNATENSGRGRSKMSTI 77
 OY 153 PKKST-----MKKATPKASPKKVAFTYINPEIHHPDKRV-----EEDDSQOK 196
 DB 78 SNESINLGLLRVNSELEESPAVHOERIKNSVANGALGHANSPKVLNKKMAQDDDKA 137
 OY 197 EDSVEPPLIOHOMKDPSCNYSDEDTNNAVPTPLHTTKPTFAQLLNKKNNEVSEPEAL 256
 DB 138 RBEKRPVKS---SSPLKTLKSTOPLSLYPES-PIH--RSSEIETNTDDEDEEDDAY 191
 OY 257 TDM-----KIKRENFSLDEKVNLYLSPTNNNSKN-VSDMDSHLQNLQDASKNK 307
 DB 192 TGLTSPOLHSRIPITNAVSIKINLDEFNLNPNESKSLVSD-----TSVDSGREL 246
 OY 308 TNEINHLNLSFALKA-----PKNDIENPLNLTNADISLSSGSSOSLOSIRNDNVLES 362
 DB 247 DKKTIPELFCMSSTPEMTVPDEKCNLPKLLNFT--SNNSHSDSRSPASV-EDLNI 303

OY 363 VEGSPKVPNPGSLINDGKGFSEVEVSLPRLDSRKLLETTKEHDAPPHNNEFDASKS 422
 DB 304 LFGADS-----SQNNPVTADALINDVNDLQON-----MEHIDAFDEKKVYDEBC 352
 OY 423 TN---TNKQOLLVSSDDHLDSDFRSYNHTEQSLINLINSASQSISLMALEKORQOE 479
 DB 353 SNEPVFLGE-----NDTRSIYYS--NKGTNAVY-----QEFSQED 386
 OY 480 QYQAAPREETSQNIKKQKPKSLNLEVKYTIKPEYSAPEIKAPKREFSSIRLRIN 539
 DB 387 SLAHSEPKFKDLNATISDDVWNEDKETDANISYTSKSESYIADYKVTROE----- 436
 OY 540 EDEIAEPADIHKKENEAHSVEDTDALKKALNDEESDTQNSTKMSIRFHIDDKL 599
 DB 437 -----DMDTKKLQSEBANEQAIIPQKDSSETELTENNESEFQNFQDGEYRI 488
 OY 600 ---EDSNDGREDND-----ISREKSDIINDYSQIS---DIIGKYGSSSE 642
 DB 489 VOHEESLYQRTKSPENIINGSEIGVDHGEAEVNEPLAKTSAEBHDLSSECEQSVSE 548
 OY 643 ITTKTLAPRSDNDKNSKSLDEPRANNESLQOQLEVPHTKEDDSLANSNTIAP---- 698
 DB 549 ARNKRIEKEVEETKDENETEKDESEYHKVEENEDEHPV-----LLPLPRW 597
 OY 699 EELTL--PVEANDYSSFN-DVTKTPDAYSSFEESLSREHETSKPINFISWHKQOK 755
 DB 598 EELQFNEPFIIDENDTSDSIDLTRSM-----KSDYSIWHIDEEI 639
 OY 756 KIQIHVPKQIITASYQYKNEQESRYTSDKV--KIPNALQFKFKEVNYSRRVSPDM 813
 DB 640 KSN-----SPESIANSQ--FSQOOSITTAIVDSKDKNGSTSPKPRIVRSRYNPKS 692
 OY 814 DDLNVSQPLPELSEDSGFDLNFANSNNMTNRRSTPLSTKVLVINDND-----P 865
 DB 693 RVSSLNYDNE-----DYLLNSEWMLDMRRNRTLSKRIQDITFQKCHAPLIR 744
 OY 866 NV-----VEPPEKSYAEIRNARRLSANKAPNOAPPLPPOQPSSTRSNS 911
 DB 745 SLMKLNGEDSGQNHFLFEBQPOEHENIPLSTLSQDITTVNG--LDQCKLPTNODEA 802
 OY 912 NKRVSFRVPTFEIRRTSSALAPCDMYNDFDDFGAGSKPTTAKGCKTLPMDKQDKR 971
 DB 803 E-----ISIREIESA-----GDI--TFNGDILLST-----SFEDELQOD 834
 OY 972 ILNAKKGYQDEYINKKLVQDKPKKNSIYVDPDRREBELQGTASINATIDSSIYGPDS 1031
 DB 835 FANFLDALDHD-----STSFNHPGDDSSSFQDSS 864
 OY 1032 -ISTDMIPYLDELAKKRP-----ALLSADRLFMEQEVNPLRNSVYLHPGAG 1078
 DB 865 KKSFNLSMSESYELKPPPSIRKQPIAPDVLOKLLKLESDTK--DQADLEKIREERT--TERPTG 922
 OY 1079 AATNSSMLPEPDFELINSFARNVSNNSDVAISGNAS--TISFNOLD-----MNFDOAT 1131
 DB 923 LGIG-----MLKTPYKQVDS-----IALASIKIGYASFSDDDSRREGMNSDAIT 967
 OY 1132 IGOKIODEPASKSANVVRGDDGLASAPETPRPTKKESSIRPAKLSASPSPKPTK-- 1189
 DB 968 LNN-----FDDFEDDKMTPSPVPR--ST-----SPTRKH 994
 OY 1190 IGSFVAVIKKNGSIAIEPIPKATHKPKKSFQNEISNHHVR--DQGISPSSGSEH--- 1243
 DB 995 VSSPFKVVK-----AANKQENNININIKAEIEIEIPMQOETDGL 1032
 OY 1244 QQHNPSSVYSQYTD--TSTVPDEKDVQHKPREKQKQKHHHHHHHHHKKQKTDIPG 1300
 DB 1033 KQDIPPLAQTKQDNVAKETITQLEPOD----- 1062
 OY 1301 VVDEIIPDGLQGRGLFVFLGIKINILPDINTHKGRTLLTDNGVCHVATTPREYVND--D 1359
 DB 1063 -VEQEPDM-----GILYLSIKATSLALYGTSHRATYAIIVDNGENVVQOTWESLPTD 1116
 OY 1360 HNAVIGKEEFLTV-----ADSL-EFILLKASYEKPROLIVEVTEK--KVVK 1403

```

Db 1117 GNIRLNKEFELPIDRKGAEIASSASSERISYKCVITLCKEKEPRHELVETIVDPVKG 1176
QY 1404 SNNRLSRLESGKDIITTTKFEVPEYK-DTWANKFAPDGSFARCYIDL-QOEFQDITGKAS 1461
Db 1177 S-----FEGKTKYKFEKKYQKKRQDEMDYLFADQDSFACELIEINEFLKNAFNIS 1230
QY 1462 QPDLNCFEMEMTMS--NGNC---PMKGRPKYKIAOLEVKMLVPRSPDREILPTPSIRSA 1515
Db 1231 HHMYIMIKMSRIADKIKGSKRLYELPRKAPHKVASLDVEAFLERTSAFEQPKQFSIV 1290
QY 1516 YESINELNNEONNVEGYLHOGDCP-IFKKRFFKMGTSLLASHSEISHKTRAKINLSK 1574
Db 1291 NKIVSKYKLOQNIYKEGYLDDGGLCKKIEKRFKLGHSQSLSGYHEISRKAKIDINLKL 1350
QY 1575 VVDLIYDKENIDRSNHNFSVDLLDHAFKIKFANGELIDFCAPKHEMKI-WIQLNQE 1633
Db 1351 VTKVLRNEDIQADNGCGRNFTDWFVFNCGFQVLPDDGERITFNACNSDEKSDWYNKQIE 1410
QY 1634 IYRRFRFRQFVNLMLQOQQQQQQQQSS 1662
Db 1411 VVELNVF-HQPWVKKYCKELAEKEKTRTT 1438

RESULT 3
MAPB-RAT STANDARD: PRT: 2459 AA.
AC P15205; Q62958; Q9ER21; Q9QW92;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [contains: MAP1
light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116.
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Testis;
RX MEDLINE=96257242; PubMed=666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
associated protein (MAP1B)-encoding cDNA.";
Gene 172:307-308(1996).
[2]
RC SEQUENCE OF 96-2459 FROM N.A.; DOMAIN, AND INDUCTION.
RX STRAIN-Sprague-Dawley; TISSUE-Brain, and Glial tumor;
MEDLINE=92347374; PubMed=1639092;
RA Zauner W., Kratz J., Staunton J., Felick P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
recombinant rat MAP 1B.";
J. Eur. J. Cell Biol. 57:66-74(1992).
[3]
RN [1]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE-Spinal cord;
RX MEDLINE=90059871; PubMed=2551510;
RA Rientz A., Gremingloh G., Hermans-Borgmeyer I., Kirsch J.,
RA Litzner U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Neuraxin, a novel putative structural protein of the rat central
nervous system that is immunologically related to microtubule-
associated protein 5.";
EMBO J. 8:2879-2888(1989).
[4]
RN [1]
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX MEDLINE=97405699; PubMed=9260743;
RA Ma D., Notthias F., Boyne L.J., Fischer I.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
in rat CNS and PNS during development.";
J. Neurosci. Res. 49:319-332(1997).
CC -1- FUNCTION: The function of brain MAPs is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes

```

```

CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -1- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC heart or muscle.
CC -1- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC nerve levels are high early in development but decrease during
CC postnatal development and are low in adults. In dorsal root
CC ganglia levels remain high throughout development.
CC -1- INDUCTION: By nerve growth factor.
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (by similarity).
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: TO MAP1A.
CC -1- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC 2459) was originally described as neuraxin in Ref.3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U52950; AAB17068.1; -.
DR EMBL: X60370; CAC16162.1; -.
DR EMBL: X16623; CAA34620.1; ALT-SEQ.
DR PIR: S06017; S06017.
DR InterPro: IPR000102; MAP1B-neuraxin.
DR Pfam: PF00414; MAP1B-neuraxin; 10.
DR PROSITE: PS00230; MAP1B-NEURAXIN; 8.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2459
FT REPEAT 1869 1885 MAP1B 1.
FT REPEAT 1886 1902 MAP1B 2.
FT REPEAT 1903 1919 MAP1B 3.
FT REPEAT 1920 1936 MAP1B 4.
FT REPEAT 1937 1953 MAP1B 5.
FT REPEAT 1954 1970 MAP1B 6.
FT REPEAT 1988 2004 MAP1B 7.
FT REPEAT 2005 2021 MAP1B 8.
FT REPEAT 2022 2038 MAP1B 9.
FT REPEAT 2039 2055 MAP1B 10.
FT DOMAIN 559 1035 GLU-RICH.
FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT KKEE AND KKEI/V REPEATS).
FT DOMAIN 2224 2312 LYS-RICH.
FT CONFLICT 127 127 M->V (IN REF. 1).
FT CONFLICT 140 140 T->S (IN REF. 1).
FT CONFLICT 2112 2112 R->K (IN REF. 3).
FT CONFLICT 2169 2169 L->I (IN REF. 3).
SO SEQUENCE 2459 AA; 269497 MW; 2E3F6872DDB8BA2 CRC64;

```

```

Query Match 4.0%; Score 342; DB 1; Length 2459;
Best Local Similarity 17.8%; Pred. No. 3,2e-06;
Matches 349; Conservative 313; Mismatches 65; Indels 608; Gaps 92;

```

```

QY 2 NSTPSKLLP-IDKSHSLQ-L--QPOSS---SASIFNSPTKPLNF-PRTNKSPSLDPNSSS 53
Db 491 NSTQVNIIEGLEKIKHLDFLKQPLATQKDLNGQVSTPVKQVKKLQKQDRESLEKLP--AT 548
QY 54 DTVTSQDQEKGEKKKPTAQTSDRNFQDLNDSIDIQTTQHQQQDQFQQQQQQLSQTDNN 113

```

Db 549 KPLSSKSVKESKEEAPETAKASOYEKT-----PKVESKEKVIYKKDKPGK----- 594
QY 114 LIDESSFOQPMSTLDLTFQONPTVKNVKNHNAPTYINTSPNKS-IMKKATPKASPKVAF 172
Db 595 -----VSKPS--VTEKEVPSKEQSPVAEVAEKATEKSPKVTYD 634
QY 173 TVTNPEIHNYPNRVEEEOQO--OKEDSV-----EPPLIOHMKPQSOFNYSDE--DTN 223
Db 635 KVVKEIKETPKREEKKEKKEVAKKEDTPLKDKPKKEEKKIKKEIKKEEKKELK 694
QY 224 ASVPRPLHTTKPTFAOLLKNNEVNSPEALTDKMLRENFNSLSDKEVNYLSPTN 283
Db 695 KEVKKETPLDKA-----KEVKDEKKEVKEKEPKKEIKKISKDKIKKSTPLSDK 746
QY 284 NNNS--KNVSDMSHQ-----NODASK-----KTNNTNIN-----LSF 317
Db 747 KPAALKPKYAKKEEPTKKEPIAAGKLKDKGVKVIKEKGTTEAATAVGTAAVAAGV 806
QY 318 ALKAPKNDIENPLNLTNDISLSSGSS-----OSSLOSLKNDNR 358
Db 807 AASGPAKLE-----AERSLMSSPEDLTKEFEELKAELIDVADIKPOLELIEDEK 858
QY 359 VLESVP-----GSPKPVNPGSLNDGIGKFSDEVESLIP-----RDLNR 398
Db 859 LKETEPEGAAYIOKETEVSAGSAESPDEGITTEG--EGECEOTPEELEPERKOGVDDIEK 917
QY 399 -----DKLETTKHDARPEHNENFIDAKSTNTKKGOLLVSSDHLDS 440
Db 918 FEDEGAFEESEBAGDYERKAE--TEEAERPEEDEDNVSGSASKSHR--TEDELIK 971
QY 441 PDRSYHTQSIILNLNSASOSQISLN-ALEKOROTEOBOTAOAPEETSQDNIKVY 499
Db 972 AEAADV-HIKEKRESVASGDDRAEDMDALEK-----GEAPOSEEEDERKADAREED 1026
QY 500 QEP-KSNLE-FYKVITIK-----EPVSAETIKAPKREFFSRILRIKNEDE 542
Db 1027 HEEDKTEADYVMAVVDKAAEAGVTEQYDFLGTAPKOPVOVSREPASSI-----HDE 1081
QY 543 IAEPAIHKKEMEANSHEEDTALLKALINDEESDTQNSIKMSIRFIID----- 594
Db 1082 -----TLPGSESEATASDEENR-----DOPEETATSGYOSTIEISSEPTPMDEMS 1130
QY 595 -----SDMKLEDSNDGREDND-----DISRFEKS-----DILANDVQSOTDILIG 634
Db 1131 TPADVMTDETNNMETESPQSEFVNITKYESSLXSOEYKRVVASPGLSGSTDTADTGR 1190
QY 635 KIGNSSEITTTKTLPAPRSDNDKENSLSLEDPANNEILO-----QOLEVPHTEKEDSILA 690
Db 1191 DYNASAS-----TISPPSSMEEDKFSKALRDAYRPEETDVKTGAELIDKIDVADERLSPA 1245
QY 691 NSNNIAP-----PEELTLPYVE-----ANDYSSFNVDYKTFDAYSSEFESLSREH--E 736
Db 1246 KSPSLSPSPSPSPEKT-PLGERSVNFSLPNEIKASAGEATVAVSPGVTAQVAEEHCAS 1304
QY 737 TDSKPINFISIMHOKOKKHQIHKVPKQIIAS-----YQOYKNEOS-----RVTSOK 786
Db 1305 PEKTELEVVS-----PSQSVTGSAGHTPRYQOSTDEKSHLLEVTYENA 1348
QY 787 VKIPNALIQFKKEKENVNMSRRVSPMDM-----LNVSOFLPELSDSGFKDL 834
Db 1349 QAVPVAFEEFTEAKDEN--ERKSISIP-MDEFVDPSESPIEKVLSPLSPLIGSESAVED- 1404
QY 835 NFANYNSNNTNRPSFTPLSKKNVLSINDPNVYPERPEKSYAELIRARLSANKAAPNQ 894
Db 1405 -FLSADDKALGRSESPFECKNKQGFSDSESVDLTSDLYDKOEKBRAG----- 1455
QY 895 APLEPQOPSSSTRNSNKNVSRFVPTFEIRRTSSALAPCDMYNDIFPD--FGAGSKPT 952
Db 1456 FIPIKEDFSPEKKASDA-----EIMSQSALA-----LDRKRLCGDGSPT 1495
QY 953 IKAEGMKTLPSMD-----KDDVYKRLNKKKGVTODEYINAKLVQDKPKRNSIVTDPED 1005
Db 1496 -----QYDVSOFGSFKEDTK--MSISEGTVSOK--SAMPVDE-----GAED 1532

QY 1006 RVELOOTASIHATIDSSIYGRPDSISTDMLPYLSDELKP-----PTALLSADR 1056
Db 1533 TYSHMEGVASVSTASVATSSF--PEPTTDVSLSLAEVGSFSTVEDDSLVSIVQTP 1590
QY 1057 LFMEOEVHPLRSNVLVHFGAGATNNSMLPEPDFELINSPARN--VSNNSDNVAISGNA 1114
Db 1591 TPQETEMSPSKKE-----CPRPMSISPPDFSPKIAKSTPRVQDHRSESSMS--- 1637
QY 1115 STISFNO-----LDKNFD-----DQNTIG----- 1133
Db 1638 --TEFQOSEPHSLAMDFFSROSPDHPVTGAGMLITENGTEVDYSPSDIODSSLHKIP 1695
QY 1134 -----OKIOEPASANTVRDODDGLASAPENPRPTKESS-- 1172
Db 1696 PTPEPSTIODNDLSELISVSQVAPSTJSANT-----PSOIASPLOEDTLSDV 1744
QY 1173 -----SKPAKLSASAPRSPKIKGSPVRYVTKKNGSIAGIE 1207
Db 1745 VPPRDMSTVASLASEKVQSLSEKLSPKSDISPLTPRES-----SPYSPGFSDSISGAK 1799
QY 1208 -----PIKATIKP--KKSFGCNELSNKVRDGGISPPS--GSEHOQHPSPW 1251
Db 1800 ESTAAAYQTSSSPPIIDAAAEPRYGFSSMLFTMQHNLALSRLDTSSVEKDNCGKTPCDE 1859
QY 1252 SVPSQYTDATSTVRDENKVOHKKRREKOKHNNHNNHNNHKKOTDIPGVVD----- 1304
Db 1860 NVAQKPESTTESPDE--EDVDYSHKETTQAHNVGGIYKTEKRT--IKSPCDSGYETI 1917
QY 1305 -----EIPRVG-----LOERKLEFRVLGIKINILPDIN-----THKGRFTL- 1341
Db 1918 EKTTPREDOGYSCELTETKTRPREGGYSYEL--SEKTTTPREVSGTYEKTERRSLD 1976
QY 1342 TLONGVCHVTTPRYNMDDHNAVIGKPELTADVADSLERITLLKASYEKPRSTIYEVYKKV 1401
Db 1977 DISNGVDTEDEGGHTLGDGSYSYETTEKIT-----SPRESESYE-----TTTKT 2022
QY 1402 VKSRNRLSRLEFGSKDITITTKFVPTVEKDTWANKFAPD--GSFAR-----CYIDLOQFE 1453
Db 2023 TRSPDTSAYETEMETIKTKTPQASTYSYETSDKCYTERKSPBEANODVCLVSSCEFK 2082
QY 1454 DQITGKASQF-----DLNCFNEMETMSNGNOPMR--GPKYIAOLEVKMLVPRSDPR- 1505
Db 2083 HPKTELSPPFINPNLEMFAGEPTESERPLTQSGAP-----PPSGGKOQ 2129
QY 1506 -----ELPTISISAYESIMELNNEQNNYFEGLIHQGGCCPLFKKFKFLMTSLLAHS 1560
Db 2130 GROCDETPPTVSSEASPSQTDSD-----VPRPETECP-----SITADA 2167
QY 1561 EISHKTRAK--INLSKVVDLIYVKENT--DRSNHRNFSVLLLD 1601
Db 2168 NLDSDESETTPDKTYVTKHMDPPRAPMODRSPRPNHVSXVD 2212

RESULT 4
YBET YEAST
ID YBET YEAST STANDARD; PRT; 1381 AA.
AC P34216;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 150.8 kDa protein in SEC17-QCRL intergenic region.
GN YBL047C OR YBL0520 OR YBL0501.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN (1)
RP SEQUENCE OF 1-961 FROM N.A.
RC STRAIN=5286C;
RX MEDLINE=95176707; PubMed=7871888;
RA de Wergiffoes P., Jacques B., Joniaux J.-L., Purnelle B., Skala J.,
Goffeau A.;

QY	208	QWMDP	-SQFVYSEDPTNAPRPPRLPHTRKPF	-AQLLNKNVNSPEALTDWK	-----	260	
Db	722	EKEEPKKEIK	ISKDKIKS	---TPQSDTKRPSALKRPVAKKEESTRKPERLAAGIKLQK	778		
QY	261	-----LKRNFNS	-----	-----LSLDEKVNLYLSPNTNNNSKN	289		
Db	779	VKVYKKKGKTE	AAATVAGTAATTAAVAAAGIAASGPVKELAEBSLMSSEDLTRKE	838			
QY	290	-----VSMDSHLQNLQDASKRKTNE	-NIHNSFALKAPKNDIENPLNSLTNADI	338			
Db	839	ELKAEIDVAKDIP	KELEIDEKELKETOPGPAVYIOMETEVSGSAESPQEGITTTT	-897			
QY	339	SLSSSGSGSSQSLDRNDRVLESVSGSKKRN	-----GLSLNDIKGSPD	---VWSL	391		
Db	898	-----GEGCEQ	-----TPEELEPEKQGV	---DDIEKFEDEGAGFEBS	934		
QY	392	LPRLSRDKLETT	KENDAPHNNEFIADKSNPTNKGQLVSSDDHLSDFRSTYHNTQOS	451			
Db	935	ETGDY	-EERKE	-TEEAPEEPDQD	---DNAGSASAKHSTEDDESAKABDVHLKERES	989	
QY	452	ILNLINSASQSL	SLNALEKQROTQEOQTOAAPEERTSFSDNTIKVQOP	-KSNLE	-F	508	
Db	990	VVSQDDRAEEDMD	-DYLEKGEAEGQEE	-----EGGEEDKADAREGEYEPQKTEADQY	1042	450	
QY	509	VKVTIKKEPVSAE	-----	-----IKAPKEFSSILRIKNEDEIAEPADIIHPKE	554		
Db	1043	MAVADKAAEAGV	TEBOYGLGTSAKOPGLOSPESSREPASSI	-----HDE	-----TLPGSE	1092	50
QY	555	NEASHVEDDALLKAL	NDEESDTQNS	-----TKMSIRFLHDSQWK	598		
Db	1093	SEATASQEN	-----REDQPEETATSGYTOSTIEISSEPTPMDEKSTPRDQWSET	1144	40		
QY	599	LEDSNDQDRDNDNDIS	KFEKS	-----DILNDVQSOTSDIIGDKYGNSSSEIT	644		
Db	1145	NNEETSPQGEFVNIT	KYESLSLQGEYSKPAVASFNGLSESGKTDATQKQYNAAS	---1201			
QY	645	TKTLAPRSNNDKEN	SKSLIEDPANNESIQ	-----QOLEVPHTRKEDDSLANSNTAP	---697		
Db	1202	-----TISPPSSMEKE	FKSALRDAYCSEEBKELKASALDITKVDSDERLSPAKSPSLSPSP	1259			
QY	698	-PELTLPVYEANDYSF	-NDVTKRFDAYS	-----SPEBSLSRKH	-ETQSKPINFISI	747	
Db	1260	SPIKKT	-PLBERSVNSLTNETIKVASAGEBARVSFGVQIQAVEEHCHASPEKLTLEVVS	-1317			
QY	748	WHQOKOKKHQIH	KVPTKQIIAS	-----YQOYKNEQES	-----RYTSQDKVRIKPAIQERK	797	
Db	1318	-----	PSQVTSAGAGHTRYQSPQSTDEKSSHLPTVESENAQAVPVFERSE	1362			
QY	798	FKEYNVMSRRVSP	EDMD	-----LNVSOPLPELSEDSGFKDLNFANYSNNNTNR	845		
Db	1363	AKDEN	-ERASLSP	-MDEVPDSESEVEKULSPLRSPLGSESEPED	-FLSADSKYLG	1417	
QY	846	PRSTPLSTNANV	SLNIDNDNVVEPEPKSYAIRNARLANKAPNQAPLP	-----899			
Db	1418	RKSSSPEEGKNGQF	-----	-----PDRESVSDLTSTGLYQDKQEKSTGTFIPLKEDG	1468		
QY	900	POROPSSSTRSNSKRV	SRRFVPTFEIRRTSSALAPCDMNDIJDDEGAGSKPTIKAEQK	959			
Db	1469	PEKTSOYEHMSSQAL	-----	-----ALDEKKLQGDVSPFOIDVQSGSP	-----1505		
QY	960	TLPMSDKDVYKRI	LNAKKGVTODEYINAKLVQDKPKKNSIYDPEDERYEELQOTASIHNA	1019			
Db	1510	-----KEDTK	-MSISEGTVSDK	-SATPVDEGV	-----ADQTSYSHMEGVASVSTA	1551	
QY	1020	TIDSSTYGRDPSI	STDMPLPLSDEKKP	-----PRALLSADLFLMEQEVHPLRSN	-106		
Db	1552	SVATSSR	-PEPTTDDVSPSLAHEVSGPHSTEVDSLSVSVQTPPTTFOETMSPSKBEC	1605			
QY	1070	-----SVLWHPGAGATNS	SMPL	-----EPFELIN	1095		
Db	1610	PRPMSISGRPDSP	RTAKSRPPYVDHNSQESMSIEFGQSPRESHFAMDSRQSRPHPLTG	1666			
QY	1096	SPARNVSNSD	-----	-----NVALISGNASTI	-----SFNOLDMNFDOATIGOKIOEOP	1141	

```

Db 1670 ASVLHTTEKPTVDVSPCDIDSSSHKTPPEEPSYTO-DNDLSLISVSO-VEASPS 1727
Oy 1142 SXSANTVGGDDGLASAPETPTPTKESIS-----SK 1174
Db 1728 TSSAHT-----PSQIASPLQEDLSDVPPREMSLVASLSEKVSLEGKLSF 1776
Oy 1175 PAKLSASPRKSPKIKGSP-----VAVIKNGS---IAGIEPLPKATHPK---KKSPOCN 1223
Db 1777 KSDISPLTPRESS-PLYSBGFSDSTSAKETAAHQASSPPIDATAETAEYCFRSMLEFD 1835
Oy 1224 EISNHNVRGGISPS--GSEHOHNPSMVSPQYTDATSTVDEKDKVQHKREKQOK 1282
Db 1836 TMOHNLALNRDLTSSVEKDSGKTDGFNYAVQKPEMAAGSPDE-EDYDYESQEXTIPI 1894
Oy 1283 NHHNNHHNNHKKOTDIPGVVD-----EIPDVG-----LQERGLKLF 1319
Db 1895 HDVVRYYUEKTEBT-IKSPDCSGSYETIEKTKTPREDGYCEIEKTKTPREBEGYCY 1953
Oy 1320 RVLGITINILPDIN-----THKGRFL-TLDNGVHCYTPPYNMDHNAVIGKEFEFT- 1371
Db 1954 EI-SEKTPTRPEVSGYUYEKTERSRRLDDISNGYDDEGHTLQDCSYSEYETTEKITS 2012
Oy 1372 --VADSLFEITLTKASYEKPRGTLVEVEKTKVKSNNRLSRLFGSKD-IITTKFVPTEV 1428
Db 2013 FESESESYETSTKTRSDPTSAYCETMEKIKTPQASTYETSDRCYTTTEKSPSEA 2072
Oy 1429 KDTWANKFAPDGSFARCYIDLOFEDOTIGKASQF---DLNCFNEMETMSNGNQPKR- 1483
Db 2073 RQ-----DVLCLVSCSECFKHPKTELSPSFINPMPLEMFAGEEPTSESEKLTQS 2122
Oy 1484 -GKRYKINOLEKMLVPRSDR-----ELIPTSIRKATESINLNEONNYPFGYH 1535
Db 2123 GDAF-----PRSGKQOGRQODETPPTSVSESAPSQNDSD-----TP 2159
Oy 1536 QEGSCDPIKRRFKMLGSLSLANSEISHKTRAK-INLSKVVDLYVQKNI---DRSNH 1591
Db 2160 RTEBEP-----SITADANIDSEDESEFTPTKTYTKHMDPRPARMDRSPS 2207
Oy 1592 RNFSVLLLD 1601
Db 2208 RHRDVSMD 2217

```

RESULT 6

```

67_YEAST STANDARD: PRT: 1658 AA.
YMG7_YEAST Q03661; Q04988;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
01-NOV-1997 (Rel. 35, Last annotation update)
Hypothetical 187.1 kDa protein in GDA1-ERG8 Intergenic region.
YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-711 FROM N.A.
RC STRAIN=S288C / AB972;
RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 608-1648 FROM N.A.
RC STRAIN=S288C / AB972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: 249809; CAAB9934.1; -
DR EMBL: 249939; CAA90180.1; -
DR SGD: S0004832; YMR219W.
KW Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;
Query Match 3 6%; Score 313; DB 1; Length 1658;
Best Local Similarity 18.4%; Pred. No. 2.8e-05;
Matches 233; Conservative 198; Mismatches 498; Indels 424; Gaps 57;
Oy 42 NSKPSLDPNSSSDTYSE--ODEKGEKKKPTAFSTFDRNF----- 82
Db 440 NTSSESDQGFADATYKKNVVEQDEDEPEKDDITRSSLDKFNHGNKNSSEYENLENE 499
Oy 83 -----DL-----DNSIDI--QOTIOHQOOQ---PQOO 104
Db 500 TDPATVERENQINDYEGYDVTGKSVESDLHESHPDLVLAARALQFOQRNSNCPQKE 559
Oy 105 QOLSQF-----DNNL-----IDFSQTPMTSTLDLTQKNPYDKYENHAPTYINTSPN 154
Db 560 EOVSSELYGHSNOSNLGSRSLDESEBQIPK--DETGNNNNLKTDRGLDSSVEIEVE 616
Oy 155 KSIWK--ATPK-ASPKVAFVTNPETINHYPDNVEEEDSQQKEDSEVPPLOHQWK 210
Db 617 KYSEKKLQDSTEKELVPLSTDTTINNSSLGNEDSIYSLDADAISENLTDVPLMEIKTT 676
Oy 211 DP-----SOFNYSQD--EDTNASVPPPTPLHTKPTFAOLLKKNNE-----VNSEPE 254
Db 677 PRKEVVISSEYVSSSTSYEDNTVAMP--POVEYTSFPMNDFNSLNDYEKKHLLKSTLA 734
Oy 255 ALTDMLKLR--NFSNLSLDEKVNLYLSPTN-----NSKNVSDMDSHLQNLQDASKNKT 308
Db 735 ALAPAFATKDAEFVEAGVATKSCLTSGHTNIFHRSKETQVSDLESTENFTFENETG 794
Oy 309 NENIHN-----LSFALKAPKNDIENPLNSLTHADISLRS 342
Db 795 DENKNQSKNFGVANSTOKSTEDNTDEKIFYSAINTVNTGDSGCCDIETATNVEENLRY 854
Oy 343 SSSOSSLSLQNDNRVLESVSPKKNVPGSLNDGKIGFSEDEV----- 388
Db 855 CEKDNNEAEMSSGDECVKQNDGSKTOISFSTSPDNFQESDNTFEFSSTKKVYNSDLE 914
Oy 389 -ESLPRDLR--DKLETTKEHA-----PEHNENFIDKSTNTKGGOL 431
Db 915 DDESLKELTKAEVVDKLDDEESEDSEYEDVADPREPDGSGNENIV--KTKKDTLGLIV 972
Oy 432 VSSDHLDSFDRSYNHTQSIILNLSASOSQISLNALEKOROTO--EQEOTOAAEPREE 489
Db 973 EPENEKVNKV-----HEETLFE-ANVSS--SYNVQKKDMHTDVIQDEAANYEAGR 1022
Oy 490 TFSQNIKVKQPKKNLEVKVTKIKPEVSATETAPKRE-----FSSRLIRKN 539
Db 1023 KYIQTNTTEAHISITETI-----DENAIGNNMEIPERSVEKETHNEVLEFRATTIEN 1077
Oy 540 EDEIAEPADIPKKEANEANSHVEDTALIKALINDEESDTT-----ONSTRMSRPHID 594
Db 1078 TKAL-----ENNTNMH--DOVSQACSDSDRQDSDTAENKNGVSAKHNLDIYVS 1123
Oy 595 SD-----WKLESDNGDREDDND 612
Db 1124 SSEISVEPLKPEDSRNIFSSPIRVIGAVYKGVVQVDAVESFYKKLDVMDSEDDQVVD 1183
Oy 613 ISREFKSIDLN-----DVSQTSIDIIGKYYGNSSS--ITTKTLAAPPSSDNKDKENSLSL 664
Db 1184 IGDYNO-DIFNKSNTSDASVNMKSVSKERDSEDEAAILGCVTEAHNDGN--NSHYI 1240
Oy 665 E-DPANNSSLOQOUEVPTTKEDDSILANSSNAPPEELTLPVLEA-----NDYSS- 713
Db 1241 NIDPTTGAYEEDSEVPROOVKD-----KENLHSE--PLVEGLQSEOHFEKKDHSEN 1292

```

```

OY 714 -----FNDVTKTFDAYSFEESLSREHETDSKPIFNISIMWKKOKKHQIHKVPTQ 766
DB 1293 EEEFPTITGDTISA-NIHNSAPDDIKR-----OQLKNLSDLEWYSOR 1334
OY 767 IIA5YQYKNEQESRVTSDKVIIPNAIQKKEEVNVMRSRVSPD-MDDLNVSOPLPET 825
DB 1335 LLEDSRGKNGOE-----SDEVMTSRERDLJTFEKSVMKEKAGAIEDTFSELDISIQPHH 1390
OY 826 SEDSGFKDLNFANYSNNTNRPSFTPLSTKNVLSINDPNVVE-----PPEPKSYA 877
DB 1391 EED-----LDLSNNGRS-----IEELNSEPEAELEYELEIEGPTFAASS 1431
OY 878 EIRNARLSANKA---PQAPRLPPQRPQSPSTRSNKRSVFRVPT---FEIRTS 930
DB 1432 KNNDEOROGNIPSTDLPPDSKEEYTDSTPYSENITAEKSAPTSPEVETISDP 1491
OY 931 ALAPCDMYNDIPDDEGAGSKPTIKAGMKTLPSPMDKDYKRLNNAKGYTODEYINAKLV 990
DB 1492 NEVPMEINDEL-----PATTLKHXDKTNTVSLD-----DRSEHLSHDY 1531
OY 991 DQPKKNSIVTPDEDRYEELQOTASIHNAITIDSSITGRPDSISTDMLPYLSDELAKRPTA 1050
DB 1532 DNEPHDINSINIKVNEGESEEHQAVDI-----PVKV-----EVKE----- 1565
OY 1051 ILSADRLFMEOEVHPLRSNVLVHPGAGATNSMMLPEPDEFLINSRANVSNSDNVAI 1110
DB 1566 -----EQEEMR--SKSVLE-----QKPSMELINDKSSPENND----- 1598
OY 1111 SGNASTISFNQDMNFDQATIGQKIQEQPASKANTVGRDDGLASABETPR 1163
DB 1599 -----ETNREKDKTKAKKSKRRKRNYSRRKRRTTEGSAASNTKR 1639

```

RESULT 7

```

MLP1_YEAST STANDARD: PRT: 1875 AA.
ID MLP1_YEAST 002455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -!- CAUTION: REF. 2 MISQUOTES THE GENE NAME AS "MLP1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entries requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB 16 H1OLOPOSSASI-----FNSPTKPLNPFRTSKPSLDPNSS-SDTYTSEQDEKG 65
OY 402 H1OLOLETFIVLEHVPPIINSFKERTDMLNNAALLLEHTSMENAKVAKELNAKNO 461
OY 66 KEEKKDTAFQTSFERNEDLDNSID---IOQTIOHQOQOPOQO-----OQLSOTDNLLD 116
DB 462 KLVECEENDLQTLTKQRLDLCROIQYLLITNSVNSDKGLPKREIQFIQIMQEDSTIT 521
OY 117 ERSFOTPMISTL-----DLTKQNPYDKVKNENHAPYIYNISPKNSIMKATPKASPK 169
DB 522 ESDSQVVERLVEFKNIIIOLOEKNAELKVBNLA-----DKLESREKSKOSLOK 573
OY 170 VAPTVNPEIHNVDMRVEEDOSQOKEDSEVPEPLIOHQMDKPSOFNYSDEPTNAPSPT 229
DB 574 ISEETVNEAKEALITTLKSEMDLESREE-----LOKELEF-LKTSVPEMEDASYNVTI 626
OY 230 PRLHTTK-----PTFAQLLNKNE-----VNSEPALD-----MKLKRENSN 268
DB 627 KOLTEFKRDLSEGOVQDLQTRISQITRESTENMSLKEIQLDLYDSKDISIKLGEKSSR 686
OY 269 LSLDEKVNLY-----LSPITNNNSKNVSDMSHLQIDAASKKTENH----- 313
DB 687 ILAEERFKLSNLTDLTKAENDOLRRFD---LVQNTILKQDKSTHTLEEVYCSKSLK 743
OY 314 -----NLSEFALKAP---KNDEIENPLNLTNDISLRSSGSSQSSLOSIRNDRVLESV 363
DB 744 IVTELLINLKEEOKLRYHLEKNLKOELNKLSPKDSLRIMWLOTLQOKRED--LLEET 801
OY 364 PGS-PRKVNPGLSLNDGKGFSDVEVESLLPRDLSRDKLETT-KEH---DAPENHNEF- 417
DB 802 RKSCQKKI-----DELIEDA-----LSLKKKTSOKDHHITKQLEDNNSNIE 842
OY 418 -----IDAKSTYTNKGQLLVSSDDHLDSFDRSYNHT----- 448
DB 843 WYONKTEALKKDYESVYTSVDSKQTDIEKLOQYVKSLSKEKIEEDKIRLHTYNYMDETIND 902
OY 449 -----EOSILNLSASQSOISLNALEKQROTOBOEOQOAE-----PEEE 489
DB 903 DSIKLELEKSKIMLTDVYSOIKKEYKDIYETTSLOQTNSKLDSFKDPTNQTINKLTDK 962
OY 490 TSPSDNIKVQOEPRSN---LEFVKVITIKREPV-----SATEIKAPREFESSRI 534
DB 963 TSLIEDKISLKEQMFPNNNLDLQKGMKEKADFKKRISILONNNKEVAVAYSEYSKL 1022
OY 535 LRIKNE-DE---IAEPADHPKKEENANSVEDTALLKAL----- 572
DB 1023 SKIONDLDQOTIVANAOANNYEBOLQKHADVSTISLREOLHTYKGOVYTLNLSRDOL 1082
OY 573 NDEESDTQNSTKMSIRFIID-SDMKLEDSND-----GDSEDDNDISRFE 617
DB 1083 NALKREKSSVSSKESLLEOLDLSNSRIEDLSQNKLLYDOIOIYTAADKEVNNSTGPG 1142
OY 618 KSDILNDVQTSIIGDKYGNSSSEITTKVL-----APPSDNN--DKENSK 662

```

Query Match 3.6%; Score 310; DB 1; Length 1875;
 Best local Similarity 18.7%; Pred. No. 4, 2e-05;
 Matches 294; Conservative 254; Mismatches 579; Indels 444; Gaps 63;


```

Db 801 TGATTAAGAAAGIAIGPAKLEERSLMSPEDLT-----KDEELKAE-VDV--- 850
QY 281 PTNNNNKVVSMDSHLOLADASKKNTNENHNLSEALKAPKNDLENLNTNADISL 340
Db 851 -----TKDIKPOLLELDEBEKLETEPVE--AVIÖKREVTKGPRESPDECITTT 899
QY 341 RSSGSSOSSLSLRNDNRVLSEVPGSPKRVN-----GLSLDGIKGFSPD---VVESSLR 393
Db 900 BEGECEQ-----PRELEPEYKGV---DIEKFEDGAGFEESSEF 939
QY 394 RDLNRDLKLETTKHDAPENHNNENFIDAKSTNTKGLLVSDHLDSPRSYNHTEOSLL 453
Db 940 GYV--BEKAE--TEEAPEEDEDGEHVCVSAKSHSPTEDESAAKADADAYIR---EKRESVA 994
QY 454 NLNLSASOSQISLNLMEKROQOQOQOQAEPEETSFSNDIKVQOEK--SNLEFVAV 511
Db 995 SODDRAEEDMD--EALKEKABOSE--EADDEDKADAREEYEPKMAEYVMA 1047
QY 512 TTKK-----EPVATPIKAPKREFSSRLIKNEDEIAEPADHPRKENE 556
Db 1048 VYDKAAEAGAEQYGLTTPKQLOAQSPGRAPASSI-----HDE-----TLPGSESE 1097
QY 557 ANSHVEDDALKKALNDEESDTONS-----TKMSIRFHIDSPWKLE 600
Db 1098 ATASDEEN-----REDQPEFFATSGYTOSTIEISSEPPMDSESTPRDVMSETNN 1149
QY 601 DANDGREDNDISREKSDIINDVSOTSDI-----GPKYGSSEIT 645
Db 1150 EETESQSGFVNITKTESLISOEYKSPADVPPLNGFSECSKTADATGADYANASAS--- 1205
QY 646 KTLAPRSDNDKENSCLIEDPANNE--SLQOOLEVPHTEKEDSLILANS----- 693
Db 1206 -TSPSSMEEDKFSALRADACSEVKASTLIDI-----KDSIASVSSKVSPPSKSPL 1259
QY 694 NTAPEEL--TLVVEANDYSF--NDVYTFDA-----YSSFEESLSREH--ETDSKPIN 743
Db 1260 SPSPSPLEKTPLEGRSVNFSLTPNEIKVSAEAEVAVSPSEYQVEVEHCASPEDEKTE 1319
QY 744 FTSIMHKOKKOKHOKHOKPTKOILAS-----YQOKNDOEGRVSDVKIPNAT---- 793
Db 1320 VVS-----PSQSVTGSAGHTPYQOSPTDEKSSHLTEYLEKPAVVSF 1363
QY 794 QFKKEKVVMSRRVSPMDMLNSQLPELSEDSGFKDLNFAANSNNTRPRSTPLS 853
Db 1364 EFSDAKDEN--ERASVSP--MDE-----PVDSSEPI-- 1391
QY 854 TKNVLSNIDNDPVNVEPPEPKSYAETIRNARLSANKAARNQAPPLPPQRPSSSTRSNSNK 913
Db 1392 -EKVLSPLRSPPLIGESAYESF-----LSADKASGRCGAESPFEER--SGKQSPD 1440
QY 914 RVSRRFVPTFFELRRSSALAPCDMYND-----TFDDGASCKPTIKAEKTLIP 962
Db 1441 QVS-----PVSEMTSTS-----LYODKOBGKSTDPAPIKEDGQ--EKKTDVEANSSOP 1488
QY 963 SMD-----KDVKRLNAAKGYTODEYINAKLVDQPKKNSIY 1000
Db 1489 ALALDERKLGDVSPQIDVSOEGSEFKEDK--MSISEGYTQSDK--SATPVDEGV----- 1538
QY 1001 TPDEDRYEELQOTASIHNAATIDSSIGRPDSITDMLPYLSDELKRP-----PTAL 1051
Db 1539 --AEDTYSHMEGVA SVTASVATSF--PEPTDVSPLSHAEVGSHTSEVDLSVSV 1594
QY 1052 LSAADRLEFMOEYHPLASNVLVHFGAATNNSMLPEPDELINSPARN--VSNNSDNVA 1109
Db 1595 VOTPTTFQETEMSPKSE-----CPRPMSISPPDPSPKTAISRTPVQDRSSQSS 1644
QY 1110 ISGNASTISFNQ-----LDMNFD-----DOATIGQKI---QEPASKSAANTVRCDDGL 1155
Db 1645 MS-----IEFGQSPQSLAMDERSQSPDHPYTGAGVLHTENGPREVODSPDMQDSSL 1699
QY 1156 A-----SAPETPRPTKESISSKPAKUS--SASPRKSPIKIGSPVR----- 1195
Db 1700 SHKIPMEBPSYQDNDLSELISVQVEASPSSTASATPSQIASPLQEDTLSDVAPPRDM 1759

```

```

QY 1196 -----VIKNGSIAGIEPIKATPHKPKKSFQGNESNHKVRD--GGISPSGSEHOQN 1247
Db 1760 SLVASTLSEKVOVSLBEGKLSPKSDISPLTPRESSPLSPFFSOSTSAVAKKTATCHSSSS 1819
QY 1248 PSM-----VAVPSQYTDATSTVPDENKDVOKPREKO 1279
Db 1820 PPIDASAEYGFGRASVLPFTMQHLNLRDLSTPLGLEKSGSKTPODFSYAOKPEET 1879
QY 1280 KÖKHHHHHHHHKÖHT--DIFGVYDDEI-----PDVGLQ-----ERKGLF 1318
Db 1880 RSPDEEYDEYSEYKTRTSDVGYYEKEIERTKSPSDSGSYETIGTKTTPEDGDYS 1939
QY 1319 FRV-----LGIKNINLPDIN-----THKGRFTL-----TLONGV 1348
Db 1940 YEIIETKTRPEEGSYSDISEKTTSPPEVSGSYEKTEKRSRLDIDISNGYDSEDEG 1999
QY 1349 CVTTPPEYMDHNAVIGKEFELTVADSLFELTLTKASYEKPRGTLVEYTERKVKVSRNRL 1408
Db 2000 TLGDPSYSYE-----TTEKITSPESEEGSYETSNTKTRPTDTSYVCYETAKEKITRTPQAS 2055
QY 1409 SRLFGSKDI--ITTKRYPTVEYKDTWANKFAPDGSFARCYIDLQOFEQITGKASQF----- 1463
Db 2056 TVSYETSDLCYTAEKKSPEARQ-----DVDCLVSSCEYKHKPTLSPSFIPNP 2105
QY 1464 DLNCFNME-----TMSNGNOPMKRGKPYKIAOLEVKMLYVPRSDPREILPTIRS 1514
Db 2106 PLEMFASEPTESEKRLTOSGCAPPPGCK--QOGRQD-----ETPPSVSE 2152
QY 1515 AYESINELNNEONNVEGYLHOEGDCPIFKKRFKMLKSTSLANSEISHKTRAK--INLS 1573
Db 2153 SAVSQTDSD-----VPPETEBCP-----STADANIDSEDESETIPTD 2190
QY 1574 KVVDLIYVKENI-----DRSNHNFSDVLLD 1601
Db 2191 KTVTKHMDPPAPVQDRSPSPRHDPVSMWD 2221

RESULT 9
USOL_YEAST
ID USOL_YEAST STANDARD: PRT: 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intracellular protein transport protein USOL.
GN USOL OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RN SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,
RA Kendrick K.E.;
RN Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE

```

CC ER AND THE GOLGI COMPLEX.
CC
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC
CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC	EMBL	X54378	CAA38253.1	-
DR	EMBL	L03188	AAB00143.1	-
DR	EMBL	U53668	AAB66659.1	-
DR	PIR	A38455	A38455	-
DR	SGD	S0002216	USO1	-
DR	InterPro	IPR002017	Spectrin	-

KW	transports;	protein	transports;	cytoskeleton;	coiled coll.
FT	DOMAIN	1	724	GLOBULAR HEAD.	
FT	DOMAIN	725	1790	COILED COIL (POTENTIAL).	
FT	DOMAIN	465	487	CHANGED (HYPER-HYDROPHILIC).	
FT	DOMAIN	991	1790	DISPENSABLE FOR THE PROTEIN FUNCTION.	
FT	DOMAIN	1172	1786	ASP/GLU-RICH (ACIDIC).	
FT	CONFLICT	847	847	G -> E (IN REF. 2).	
FT	CONFLICT	924	924	E -> K (IN REF. 2).	
FT	CONFLICT	1253	1253	V -> I (IN REF. 2).	
FT	CONFLICT	1319	1319	I -> V (IN REF. 2).	
FT	CONFLICT	1461	1461	N -> S (IN REF. 2).	
FT	CONFLICT	1581	1581	I -> S (IN REF. 2).	
FT	CONFLICT	1600	1600	G -> V (IN REF. 2).	
FT	CONFLICT	1661	1661	R -> S (IN REF. 2).	
FT	CONFLICT	1772	1772	D -> DEEDDEE (IN REF. 2).	
SO	SEQUENCE	1790	AA: 206424	MM: 6CEB216E9FDD818	CR:664;

Query Match	3.4%	Score 292.5	DB 1	Length 1790
Best Local Similarity	17.0%	Pred. No. 0.00019		
Matches 244	Conservative 254	Mismatches 484	Indels 455	Gaps 57

[illegible]

OY	392	--PROJLSRKLLETTKEHADPAPHENNENFIDASTNTNKCOLLVSSDDHLDSPDRYN--	446
Dd	915	IKAIINENLEEKICQNNLKSKEKHEIISSELVEYYSRSROSHNLVAKLTETLKSLIANNYKRM	974
OY	447	--THEOSTLNLLNASOSQSISLNALEKO--ROTQEODOTOAAPEEBEETSFSDNIKRYOEPP	502
Dd	975	QAENESMLIKAVEESKNESSTIOLSNLOKIDSMSQENFOI---BRGSIKENIEBOLKKT	1030
OY	503	KSNLEFVAVTIKKRPVSATELKAPKREPSRLIRKINEDIAEPAD-----	548
Dd	1031	ISDLE----OTKEETISKSD--SKDEYESOILSKLELETATTANDENVNKISLTKTR	1087
OY	549	-----IHDKNEANSHVEDTALKLKALNDD-----ESDTTONSTKM-SIR	590
Dd	1085	EELAEALAAYNKLNKELETKETSEKALKEVKENEHLKREKIQLKEKATEPIKQOLNSLR	1144
OY	551	FHISDMKLEDSNGC-----DREDNDISFEKESDLINDVSOOTSIGDC	635
Dd	1145	ANLES---LEKEHEDDLAOLKYEEQJANKEROYNIENSOL--NDBITSTOOENESIKRK	1199
OY	636	YGNSSETITTTLAPPRSDNNDKENSKSLDPANNESLQOOLEVPPTKDDDSLANSNR	695
Dd	1200	-NDELCEEVKAM---KSTSEEOSNLKKSITDALNLOT-KELKKKHETWASILESIXSV	1255
OY	696	APPEELIPVEANDYSEFNDVTYTFDAYSESFEBSLSREHETDSKPINFISTWHKOEOK	755
Dd	1254	--ESEYVIKIELDDECNFFE---KEVSELDKCLKASEDKSKSYL-----ELCK	1296
OY	756	KHOJHKVPTKOIIASYOOKNEOESRVTSQVKVIPALOFRKRKE--VNMVRVRVSPDM	813
Dd	1297	ES-----EKIKEELDAKTTELKIOLEKITNLSKAKESEBELSRKKTISE	1344
OY	814	DDLNVSQFLPEPLSDESGFKDLNFANYSNNTRPRSTPYSTKANVLSINDNPVVEPPER	873
Dd	1343	ERKNAEEOLEKLKNEIOIKNOAF-----EKERRKLLNESGSTTOYESEKINTLE---	1391
OY	874	KSYAEIRNARURLSNKAAPNOAPPILPPOBPSSSTRSNSMKRVGRFVPEFIJRISALA	933
Dd	1392	DELIRLONENLAKK-----EIDNRSELE	1411
OY	934	PCDMYNDIEDDFGAGSKPTJKABGMKTLPSMD-----KDVYKR---ILMAKKGYTOD	982
Dd	1417	KVSLNSND-----ELLEEKONTIKSLODEILSYKPKITRRNEKELLSTERDOKRDL	1465
OY	983	EYINAUKLVDOCKPRKNSIVTPDEPKYEELOOTASTAHNATIDSSITYGRPDISINDMPLYSD	1045
Dd	1466	ESLBEOQLAAAESKAY---EBGLKLLBEESSKEKEALELK-----SEKMMKLLS	1511
OY	1043	ELKRPPTAL-----LSADRLFMEQEVHPLR-SENSVLVHPGGAANTSSMLPE	1086
Dd	1513	TIESNETELKSMETIRKSDPEKLEQSKSAEDDIKNLOHEKSDLI-----SRINES---E	1566
OY	1089	PDEFELINSPAR-----WVSNNSDVAISGNASTISTFNOLD-----	1122
Dd	1565	KDIEELKSKLIEAKSGSELEYKQELIMNAQOEIRINAENTVYLKSKLEDIERELKDQOA	1624
OY	1124	--WMFDQOATIGOKIOE-----OPAKSAMTVRG-----	1150
Dd	1625	EIKSNQOEKELITSRLKELEDELDSIOQKAKOSEEBERRAEVRKFQYEKQSDDEKAMLETT	1687
OY	1151	-----DDGLASAPETPTPTPK-----ESISSKPAKLSASAPRKSPI	1188
Dd	1685	KYNDLVNKEQAKMEDDETVAKTTDOSQOELEKLAKELDNKAKANSKVLKANEDEBSRT	1741
RESULT	10		
TANA_XENLA		STANDARD;	PRT; 1744 AA.
ID	TANA_XENLA		
AC	001550:		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Tanabin.		

OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tadpole head;
RX MEDLINE=92398961; PubMed=1524825;
RA Hemmati-Brivanlou A, Mann R.W., Harland R.M.:
"A protein expressed in the growth cones of embryonic vertebrate
neurons defines a new class of intermediate filament protein.";
RL Neuron 9:417-428(1992).
CC -I- TISSUE SPECIFICITY: GROWTH CONES OF EMBRYONIC VERTEBRATE NEURONS.
CC -I- DEVELOPMENTAL STAGE: IS EXPRESSED IN THE NEURONAL AND PERSIST
CC DURING EMBRYOGENESIS IN THE BRAIN, CRANIAL NERVES, AND SPINAL
CC CORD.
CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -I- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M99387, AAA4966.1; -.
DR PIR: JH0720, JH0720.
DR InterPro: IPR001664, IF.
DR Pfam: PF00038, filament; 1.
DR PROSITE: PS00226, IF; 1.
KM Intermediate filament: Coiled coil; Neurone.
FT DOMAIN 1 12 HEAD.
FT DOMAIN 13 314 ROD.
FT DOMAIN 315 1744 TAIL.
FT DOMAIN 8 48 COIL 1A.
FT DOMAIN 49 60 LINKER 1.
FT DOMAIN 61 156 COIL 1B.
FT DOMAIN 157 179 LINKER 12.
FT DOMAIN 180 193 COIL 2A.
FT DOMAIN 194 199 LINKER 2.
FT DOMAIN 200 314 COIL 2B.
SQ SEQUENCE 1744 AA; 199561 MW; 6502EAC9F6C4E93 CRC64;
Query Match 3.4%; Score 291; DB 1; Length 1744;
Best Local Similarity 18.6%; Pred. NO. 0.00021;
Matches 303; Conservative 258; Mismatches 614; Indels 458; Gaps 77;
QY 68 EKKDIAFQTSFDRNEDLNSI-----DIQOTIHOQOQPOOQOOLSTQDNLI 115
b 314 EAESRITDYRGSTYFNDMSLEHNNVRRROSEDTKRTYKSHRYSKKQIG--DKNLI 371
QY 116 DEFS---FQTPMTS-----TLDLTIKONFTVDKVENHAFYINTSPNKSIMKATPKASP 167
Db 372 QRPSTLNNSTVKSASAVPVRTSPVTKEFQVSSVLSOGILKTKARQAVEVQVSTVK-- 428
QY 168 KKAFTVNPPEIHNY-----PDMRVEEDQ-----SQOKEDSVEP 202
Db 429 -----SNLEHTTFSGAFRRRAQVETKKTDEQVIKKDALGLINDLNKNGFKEEDIQP 481
QY 203 PLIQHQMDSQFNSDSDTNASVPPPLHTTKPTFAOLLNK-----NNEVNSEPEALT 257
Db 482 GFMDH-----VKSQSVSTEKHYVEIDPLLESALKSLEDLSVSTTFNAGOSSNLEAIK 535
QY 258 DM---KIKRENFSLDEK---VNLVLSPTNNNSKNVSDMDSHLONLQDASKNKTNE 310
Db 536 DVLGPICLENONEIAFEKESPGTNAADPIEEVISESVYQYHFEK-QELSNILETE 594
QY 311 NIHNLSFLKAPKNDIENPLNLTNADISLRSSGSSQSLQLRDN-----RYLSEV 363
Db 595 NTH-----ENHVODATQAFNSCEQDGHDRAS--TLENNPEVDVQYIIRTL-- 637

QY 364 PGSPKPVNGLSLNDGKSGFDEVESSLPRDLRSKLETKTEHDAPEHNNENFIDAKST 423
Db 638 -----SNEIKESKIPSD-----NTEBAELISRSRVLE----- 666
QY 424 NTNKGOLVSPDHLDSFDRSYNHTEQSLNLNLSASQSLNMALEKORQOEQOTQA 483
Db 667 ---NEYIPVSKDDLTEFT---SHLE-----NDSQSQSPSKLFENKSTEDQLITNL 712
QY 484 APEEETSFSDNIKVKQOEKSNLEFYKV---ITKKEPVATETKAKRREPSILIRKNE 540
Db 713 KNTDENIFQSN---QEHLENLEFDSVDPVTK-----FMYPOE 748
QY 541 DEIAEPADIIHPKKE-----NEANSHEVDLALKLKNDEESDPTQNTKMSIRPH-ID 594
Db 749 NNLLEENYVGGGELVOMATDENIINQSDOLL---LSHSHHEETKTESIAVENRNE 805
QY 595 SPW-KLEDSNDGREDNDNIS---REFKSDILNDVSQF--SDIIGDKYGNSSSEITTKT 647
Db 806 SHEAEYDKSSEIPEVEISENVSEETIHEISDVEEDTKQAFEDERVGEQINONNOE--ST 862
QY 648 LAPPSDNDNKNKSLSDPAN-NESTLOQULVPHTKEDDSILANSNTA-PEELTLYV 705
Db 863 VLDGVSYSQEENSQLEDEVSISEQIEKDFEI---NEQECKSDQIREAPFTEEDVHOV 919
QY 706 VANDYSSE-----NDVYKTFDAVSSPEESLSREHETDSKPINFISIMHKOEK--OK 755
Db 920 VPFMQEQSFEREVGOGLNNIKOEVDYLONTDED---SFQNNDEPQELSCDLOEKLEEE 976
QY 756 KHIQHKVPTKQIIA-----SYQYKNEQESRVTSDKVIPAIOFKKFEVNVMSRRV 808
Db 977 ENQISENEGQNGNDIEEFQSQGYDTDEIQET-----IGNQVSAQLCESDINDQKLT 1031
QY 809 VSPMDDLNVSQLPELSDSGFKDLNFANYSNNTRPSFPTLSKVNLSIANDPNV 868
Db 1032 SMEDEEON---NPETENIGLEQ---ESDQENTRSNEGKFS-----QEECDV 1075
QY 869 EPEPEKSYAIRNARLSANKAAPNOAPPLPPORQPSRSNSKNKVSFRVPTFEIRRT 928
Db 1076 EKPED-----MSKSEYSGQGEDLDQVYDFSLNBEQANDL-----LEKE 1115
QY 929 SSALAPCOMYNDIFDDFGAGSKPTIKAEGMKTLPSMDKD---DYKRILNAK--GVYQ 981
Db 1116 EYILHHAHQDQRSVNDIEITIDEKLSERI-----IDNELATVDVNESLAANEQVDLPT 1167
QY 982 DEVI---NAKLVN---QPKKNSIYTPDEDKRE--ELQOTASINHA-----TIDSSY 1026
Db 1168 DEYAADNDYGMQDDDSGQYQTKEDLFVNDGNNTIEKIEIQOTSLNLQICEYVDNVEDIS 1227
QY 1027 GRPDSISTMLPYLSDCLKPPYALSLADRLPMEQEVNPLRSNSVUVHFGAGATSSML 1086
Db 1228 GEAKNESVEM---NDVVDLYPEAKVYTG---EQISPLQDEKILNLEKMETDKNDGDL 1278
QY 1087 -----PEPDF-BLINSF--ARVNSNSDNVAISGNASTISFNOLDNPFDDQAT--IGOK 1135
Db 1279 CLEKENETEYIEVTDSPQFATLSDH-----AGRELIVDQNSANLQFCENPFTKLIANH 1332
QY 1136 IOEQAPKASANTVRQDDGLAS---APEPPTPRPK-----KESISSKPAKL 1178
Db 1333 IEYETVA-----DSLESTEEQVOLETERTPEFPEDSKMENENSESESVDSQELSL 1383
QY 1179 SSASPRKSPKIKIGSPVRYVKKNGSIGIEPIPAKTKPKPSFGQNGISNHKVVDGISPS 1238
Db 1384 NSHKSEFE-----ISKDYQLE--QTLRVDYTLPLNLEDEFEDELTE-----QPD 1424
QY 1239 SGSEHQINPMSVPSQYDTATSTVPDENKDVOKRPREKQKNNHNNHNNHNNKQTDI 1298
Db 1425 VHEEHQNDNDGAS-----TFTTSVDEDKEREYRESVSKDEESN----- 1463
QY 1299 PCGVNDEITPDVGLQEGKLFYVLGKIKNINLNDINHKGRFTITLDNGVNCVTPPEYMD 1358
Db 1464 ---EEFEGD-----VLSDKTSQVEVTTLSG-----LAQEPSYIAD 1496
QY 1359 DHNVAIGKEFELTVADSLFEITLKASYEKPRCTIV-----EYTEKKVY- 1402


```

Db 1497 N-----ESEE-----DSHENAELN-----ENPSNDIVDFWVSOETETKIIIAEQVTEQEV 1543
Oy 1403 -----KSNRNLRLFGSSDIITTTKEVPTLEKDTWANKFAPDGSFARCYDILQOEDQIT 1457
Db 1544 TLQEDDAFNKLENNARE-KETYDESNEENIEFTNE-NQASPAANDIVDENQSEDSVI 1601
Oy 1458 GKASOFDLCNFEMETMNGNCPMKRGPKYKAOLEVKMLYVPRSDPREILFTSIRSAVE 1517
Db 1602 SD-----NEGTTSSYEDLPNATPSISHVVALESN-----STEGSSYDTKRMTE 1647
Oy 1518 --SINELNNEONN 1528
Db 1648 GYEITSLONVEDN 1660

RESULT 11
CENE_HUMAN STANDARD: PRT: 2663 AA.
002224:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RT Nature 359:536-539(1992).
RL [2]
RN CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thowar D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RT EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RT J. Cell Biol. 143:49-63(1998).
RN [4]
RP FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
RP KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
RP OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
RP AND/OR SPINDLE ELONGATION.
RN [5]
RP SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
RN [6]
RP SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
RN CONGESTION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
RN QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
RN [7]
RP SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
RN [8]
RP THIS SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation-
RN at the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN [9]
RN EMBL: Z15005; CAAT8727.1;
RN PIR: S28261; S28261.
RN HSP: P17119; 3KAR.
RN GeneW: HGNC:1856; CENPE.
RN MIM: 117143;

```

```

DR InterPro: IPR001752; kinesin_motor.
DR Pfam: PF00225; kinesin_1.
DR PRINTS: PRO0380; KINESINHEAVY.
DR SMART: SM00129; KISC. 1.
DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
DR Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN 1 335 KINESIN_MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NR_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 2663 AA; 312087 MW; CEC13880C8C8C8 CRC64;

Query Match 3.4%; Score 290.5; DB 1; Length 2663;
Best Local Similarity 17.6%; Pred. No. 0.00036;
Matches 323; Conservative 319; Mismatches 757; Indels 435; Gaps 75;

Oy 58 SEQDEKGEKKKDYAFQTSFDRNFDLNSIDIQQTIGHQOQOPOQOQOLSGTDNNL-ID 116
Db 887 NEMEDLKEQLERNDSPLQT-VEREKLITE-KLQOTLEEVKTLTQEKDQLQOESLOIE 944
Oy 117 EFSFQTPMTST----LDLTKOMPVDKVNENNAPIYINTSPKSKIMKATPKASKKAAV 172
Db 945 RDQLKSDIHDTVMNMNIDTQOELNALLESIKOHOET-INTLSK----- 986
Oy 173 TVTNPEIH---HYPDNRYEEDSOOKEDSVPEPLIQHQMKPDSQFNYSDEPTNASVPPT 229
Db 987 --ISEVGRNLHMEENTGETNDEPQKMGID-----KKQDELANQTQTLTADVADN 1036
Oy 230 PPLHTTKPTPEAOLNKN-----NEVNSEPEAL-TDMK---LKRENFNSLSL--DE--K 274
Db 1037 ELIEQQRKIFSLIOEKNELQMLSVIAEKEDLKDLENTIENIENIEBELLELDELK 1096
Oy 275 VNLVSPPTNNNSKNVSDMD---SHLOMLQASKKKTE-----NIHLSFALAKP 322
Db 1097 QOEIYAQEKNAIKKEGELSRPCDRLAEVEKAKSQQLQOQOLLNVQEMSGMK 1156
Oy 323 KNDIENPLNSLTNADISLRSSGSSQSSLOSRLNDNRVLESYVPSGR--KVNPGSLN--- 377
Db 1157 INEINLKNELKNKLETLHEHMETERLELAQKLEN--YEVKSTIKKKKVLKELQKSET 1214
Oy 378 --DGKGFSDVEVESLPRDLSRDKLETTKE---HDAPEHNNEPDAKSTNTNKGQL 431
Db 1215 ERDHLRGYI-----REIATGLQTKBEKIAHILKHEQETIDELRSVSEKTAQI 1265
Oy 432 VSSDDHDSFPRSYN-----HTEOSILNLSASOSISNALF---KQRTQ--- 476
Db 1266 INTQD---LEKSHTKLOEELPVLHEDELLPNVKKVSETQETNLELLTLEOSTTKDST 1321
Oy 477 -----EBOQTAAEBEETSFSDNIKVKQEPKSNLEFVKVTKKPYASATEIKAPKREFS 531
Db 1322 TLARIEMRLRLNEKFOGS--QEELKSLTKERDNKTIKALE---VNHDLKEIRRTL 1376
Oy 532 SRI-----LRKNEDF-----IAEPADHPR-----KENE 556
Db 1377 AKIESOSKQOSLNKKKMDNETTKIVSEMOFPKPSALRIETELMGLSKRLQESHDE 1436
Oy 557 ANSHVEDTDL--LKKALNDEESPTTNSKMSIRFIIDSDMKLESDNDQDREDNDIS 614
Db 1437 MKSYAKKEDDRLQOEVL--QSESDQLENKEIKELVAKHLEELKEEELKVAHCKLQOBEETIN 1494
Oy 615 RF-----EKSDLNDVYSQSDIIGDKYGNSSSEITTTKTLAPPRSDNNDKENSKSLDPAN 669
Db 1495 ELRVNLSKRETEISTIQOLKLAINDKLNQTOELYEK-----EEGLNKKQISEVOE 1545
Oy 670 NESIQOOLEVHTKEDDSILANSSNIAPPEELTPVVEANDYSSFNVDYTKFDAYSSFE 729
Db 1546 NVNELKQFK-EHRAKDSAL--QSIESKMLELTRLOESOE--EIQIMIKKEEMKRRQOE 1600
Oy 730 SLREHETDSKPINFISIMHROEKOKHOIKV---PTKOILISYQOYKNQES-EVTS 784
Db 1601 ALQIERQLKENTKEIYAKKKESEKEYOFLKMTAVNETQEKKEIEHLAEQFETQKLN 1660

```



```

OY 785 DKVILPNAIOFKFEVNVNSRVVSPDMDL-NVSOELP-----EL 825
DB 1661 EIEIEN-IRLTQILHEINLEEMKSVTKERDLSVEETLKVERDOLKEMLRRTITRDLEK 1719
OY 826 SEDSGFKDLNFANYSNMTNRRPSFTPLSTKNVLSNIDND-----PNVPEPEPKSYAIFRN 881
DB 1720 QELKIVNHNHLEKHQETIDKLGIYSEKT-NEISNMOKLEHNSNDLKKODLKIQELR- 1777
OY 882 ARRLSANKAARPOAPPLPPOQOPSSSTRSNKRVSEFRVPTETIRTSALAPCDMYNDI 941
DB 1778 ----IAHMLKQOEETIDLRGIVSEKTDLSNMOK-----DLEN-- 1813
OY 942 FDDFGASPTIKAGOMKT-----LPESMDKDVKRLINAKGVTODEYINAKLVDPKKN 997
DB 1814 -----SNAKLOEKIOELKANEHOLITLKD-----VNETOKVSEMEOLKKOLKODS----- 1860
OY 998 STVTPDEDEYELQOTASIH-NATIDSSITYGRDSSIS--TDMPLYLDELKRPPTALLSA 1054
DB 1861 --LTLSKLEIENMLNLAQELHEMLLEMKSVKMERDNLRAVEETLKLERDOLKESLOETKAR 1918
OY 1055 DLFMEQEVNPLRNSVLNPRGAGATNSSMLPEPPELINSFARVNSNSNDVAISGNA 1114
DB 1919 D-LEIQOELKTAR-----MLSKHEKETVDKLREKISE----- 1949
OY 1115 STISFNOLDMNDQDQATIGOKIOEOPASAMTVRGDDGLASAPETPTPTPKESISSK 1174
DB 1950 KTIQISDIOKDKSDKDELQOKIIOELQKELDOLLRKVEVNNSHKINEMEOLOKOPERN 2009
OY 1175 ---PAKLSSASPRKSPFIKIGSPRVYIKNGSIAIEPIPKATHPKPSF-----QONE 1224
DB 2010 YLCKEMDMFOJLTKHNELEIRIYAKERD--ELARIESLKMEDQFIATIREMLAND 2067
OY 1225 ISNKKVROGIGIPSSGSEHQDN-----PSMVSVPQYDATSTYPERDKDVO 1272
DB 2068 RONHVOYKRRKRLISDGOQHIMESLREKCSRIKELLYKRVSEMDHYECLRLSLDEKETE 2127
OY 1273 -HKRPKOK-----OKNNHNNH-----HHNKOTIDRGV 1302
DB 2128 FRIKAKKLVYUVTYTKIEBOHECINKFEMOFIDEVEKOKELLITQIOLQDCDVP--- 2184
OY 1303 DDEIPDVQIÖRGKLFERVLGIKINLPDINHKGRTLLDNGVHCVTTPR-----YN 1356
DB 2185 SRELRLKLNQMMDLHIEEI-LKDFSESEPSIKTEFOQVLSNRKEMTOFLEMLNTRFD 2243
OY 1357 MODHNAIÖKEFEELYADLEF-----ILTLKASIEKRGITVEYTEKKVYKSRRLSR 1410
DB 2244 IEKLNGLOKENDRICQVNNFNNRIIAINESTEFERSATISKEMEDLKSLEKENEK 2303
OY 1411 LFSGKDIITTTKFPVTEVADTANKFAPDGSFARCYIDLQOFEDQITGKASOFDLNCFNE 1470
DB 2304 LKKNVOTLKTSLASGOVAPPTQDNKP-----HTSRAOTLTTKIRE 2347
OY 1471 WETMSNGNOPMKRGKYKLAOLEVKMLYVPRSDPREILPTIRSAYESINELNN--EQNN 1528
DB 2348 LENSLSHEAKESAMHKSKEIKKQEL-----EVTNDIIAKLOAKVHESKCEKTK 2398
OY 1529 YFEGYLHOGS-----DCPIFKRFRKL-----MGTSLASLSESHKTRAKINLSKV 1576
DB 2399 ETIOYLODVVALGAKPYKEIEDLKMKLVKIDLEKMKNAKEFEKIS-ATKATVEYOK-- 2455
OY 1577 DLIYVDKENIDRSHNHFSDVL----- 1600
DB 2456 EYIRLLREULRNSQOQODTSVISEHTDPOPSNKPPLCCGGSGIVOMTKALILKSEHRL 2515
OY 1601 DHAFTIKFANGELI-----DFCAPNKH-----EMKIW 1627
DB 2516 KEISIKLQOONOLIKOKNELLSNNHLSNEVKTW 2549

```

```

AC P70478;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein).
GN APC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fischer 344/N; TISSUE=Brain;
RX MEDLINE=96116966; PubMed=8563176;
RA Toyota M., Ushijima T., Kakuchi H., Watanabe M., Imai K., Yachi A.,
RA Sugimura T., Nagao M.;
RT "CDNA Cloning of the rat APC gene and assignment to chromosome 18.";
RL Mamm. Genome 6:746-748(1995).
RN [2]
RP MUTAGENESIS.
RC STRAIN=Sprague-Dawley, and FISCHER 344/N;
RX MEDLINE=95148647; PubMed=7846077;
RA Kakuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,
RA Weisburger J.H., Sugimura T., Nagao M.;
RT "Specific 5'-GGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon
RT tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).
CC -I- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
CC CATEININ. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
CC ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATEININ (BY
CC SIMILARITY).
CC -I- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATEININS (BY
CC SIMILARITY).
CC -I- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
CC -I- SIMILARITY: CONTAINS 7 ARM REPEATS.
CC CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL: D38629; BAA07609.1; -
DR HSSP: Q02248; 3BCT.
DR InterPro: IPR000225; Armadillo.
DR Pfam: PF00514; Armadillo_seg; 6.
DR SMART: SM00185; ARM; 5.
DR PROSITE: PS50176; ARM_REPEAT; 1.
KW Anti-oncogene; Phosphorylation; Coiled coil; Repeat.
FT DOMAIN 1 728
FT DOMAIN 125 260
FT REPEAT 451 493
FT REPEAT 503 545
FT REPEAT 546 589
FT REPEAT 590 636
FT REPEAT 637 681
FT REPEAT 682 723
FT REPEAT 724 765
FT DOMAIN 739 2831
FT DOMAIN 1130 1155
FT DOMAIN 1556 1575
FT DOMAIN 1864 1891
FT MUTAGEN 523 523
SQ SEQUENCE 2842 AA; 310530 MW; 3CB32EA8A3458F47 CRC64;

```

Query Match 3.4%; Score 289.5; DB 1; Length 2842;
 Best Local Similarity 20.3%; Pred. No. 0.00043;
 Matches 295; Conservative 188; Mismatches 514; Indels 457; Gaps 68;

DR SMART; SM00526; H15; 1.
 DR SMART; SM00249; PHD; 2.
 DR PROSITE; PS01359; 2F_PHD_1; 1.
 DR PROSITE; PS0016; 2F_PHD_2; 2.
 KW Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
 KM Nuclear protein.
 FT ZN_FING 206 265 PHD-TYPE 1.
 FT ZN_FING 259 313 PHD-TYPE 2.
 FT DOMAIN 371 379 POLY-SER.
 FT ZN_FING 538 560 C2HC-TYPE.
 FT DOMAIN 788 801 POLY-GLU.
 FT DOMAIN 989 995 POLY-GLU.
 FT DOMAIN 1019 1026 POLY-ARG.
 FT DOMAIN 1069 1078 POLY-GLU.
 FT DOMAIN 1147 1150 POLY-LYS.
 FT DOMAIN 1221 1242 GLU-RICH.
 FT DOMAIN 1267 1302 GLU-RICH.
 FT DOMAIN 1411 1414 POLY-GLU.
 FT DOMAIN 1593 1597 POLY-SER.
 FT DOMAIN 1643 1704 GLN/PRO-RICH.
 FT DOMAIN 1897 1977 MET-RICH.
 FT SITE 1546 1547 BREAKPOINT FOR TRANSLATION TO FORM MOZ-CBP.
 SQ SEQUENCE 2004 AA; 225054 MW; 9FFBAC3792854BA CRC64;
 Query Match 3.3%; Score 287.5; DB 1; Length 2004;
 Best Local Similarity 18.2%; Pred. No. 0.00034;
 Matches 257; Conservative 197; Mismatches 509; Indels 449; Gaps 58;

15 SHLGLOPQSSASIFNSP--TKPLNPRTNSK---PSLDPNSSDPTVTYSEDOQEKKEE 68
 818 SHENKODSYSESEKKPEMAFVSSTRLSKVLPDLSLPANS-----QSSRRGWMG 869
 69 KQDTAFQTFEDNRFDLNDSDIDQITIOHOQO-----PQOOQOLSDTDNNLIDFST 120
 870 RNRKRTQERFG---DKDSKLLLEETSSAPQEOYEGEGEKSEATQEOYTESEQLV----- 921
 121 QTPMTSTLDLTKQNPVDVNVENHAPTYINTSPNKSIMKATP-----KASPKVAFTV 175
 922 -----ASEQPSQDKPD-----LPKRRLSEGVWRQOIKSPALCRLT 963
 176 N-----PELHHPDNRV-----EEDSOQEKEDSVPEPLIQHOKWDSQF----- 215
 964 EGSERLPRRYSEGDRAVLKGFSESESESESESESPSPILTKPKLKKKPKPLHRRRV 1023
 216 -----NYSDEDTNASVPTPLHTTKPTFQNLKKNNEVNS 251
 1024 RKRKHNSSVYETTSETTEVLDEPEDSDSER-----PMRPLEPTF-----ETDE 1069
 252 EPEALTDML-KRENFSLNLDKENVLYLPTNNNSKNVSDMSHLQNLQDASKKKTNE 310
 1070 EEEEDENLFPREYFRRLSSQDLRC-----QSSSKRSKDEDEDEDDDDPTILK 1123
 311 NTHNLSFALKAPKNDIEN-PLNSLTNADISLRSSGSSQSLQSLRNDNRVLESVPSGPKK 369
 1124 PVSLLR-----KRDVKNPSLEPDTSTPLKKKGMPKSRKRPDIHWKR-----PGR--- 1169
 370 VMPGLSLNGIKGFSD--VVEST--LPRLDSKLETTEKHAPEPHNNENFTDAKSTNT 425
 1170 -KPGFLSEIMPVSTOACVIEPIVAPKAGRPKIQESETVEYEPKED----- 1216
 426 NKGOLLVSDDHLDFDRSNYHTEQSLNLNLNSASQSLNLNLEKORQTOEQOQAAE 485
 1217 -----MLPERKEEEMQAEBAEAE 1238
 486 PEETSFSDNIKVKQPKSNLEFVKVTKIKPEVSATEIKAPKREFSRILIRINDEETAE 545
 1239 GEEEDAASSEVPASPADSS-----NSP--ETETKPEVEEERKPRVSEEQOSE 1287
 546 PADIHKKENANSHEVDLALKALNDEEDTTONSKMSIRRHIDSKMLDSNQG 605
 1288 ----EEQBLEPPEEEDAAAEATQONDHDADDEDG-----HLESTKKKLELEOP 1336

QY 606 DREDNDISFEKSDILNDVSQTSIDIIGDKYGNSSSEITTKTLAPPSDNNKDKSKLE 665
 1337 TREDEVKEFPVOSEFLDANQKREKDKD---EETLDBEEDQPSHDTSVSONMGE 1393
 QY 666 DPANNESLQOQLHVPHTKEDDSILANSSNIAPPEELTPVYEANDYSSFNVDYKTFDAYS 725
 1394 DDEEDDS-----HTKEELIELKEEBEI-PSHELDLFTVQAVQSLQESSSEHCAQY 1444
 QY 726 SFEESLS-----REHETDSKP--INFISTMHKQEKKHQIHVPTKQIIASVQOYKNEQ 778
 1445 DCEETLACQTLQSYQADEDPQMSWEDCHASE-----HNSP-----ISSVQSHPSQS 1493
 QY 779 ESHVTSQKXKIPNAIQFK--FKENVNMSRRVSPDMDLNVSOFL--PELSE-----DS 829
 1494 VREVSS-----PVPALFESQYQISPPQGLSAPSMQNMETSPMDVPSYSDHSQYVDS 1548
 QY 830 GFKDL-----NPNAYS-----NNTNPRSFPLSTKNVL----- 858
 1549 GFSDLGSIESTTENYENPSYDSTMGGSIGCNSSSQSSCSYGLSSSSSLTQSSCVYTOQ 1608
 QY 859 -----SNIDNDN-----VVE-----PPEKSTAEIRNARLS 886
 1609 MASMGSSCSMGOQSVQPAANCSTKSPQSCVVERPPSNQOQPPPPQPPPPQPP 1668
 QY 887 ANKAAPNQ-----APLPPOQOPSTRSNSKRVSRFPVTFEIRRTSSAL 932
 1669 APQPPPPQPPQPPQPPQPPQPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1726
 QY 933 APCDMYNDIFDDFGAGSKPTIKAEGMKTLPSMDKDVKRIILNKKGVYODEYINAKLVQ 992
 1727 GNISYIERIPDGFAGS-----YSQPSA-----TFSLAKL--- 1756
 QY 993 KPKKNSIVTDPEDRYEELQOTASTIHNTIDSSIIYGRDSDISTDMLPILSELKPPALL 1052
 1757 -QQLTNTIMP-----HAMPSHSPAVTS--YATSVSLNGLQALAPS----- 1797
 QY 1053 SAURLFMEQVHPLRSNLSVLVHPGAGA-ATNSSMLPEPDEL-----INSPARNVNSDN 1107
 1798 -----HPLAGTPOAATITPTPNLSTMTNLTLSPLOCMKATN 1836
 QY 1108 VAISGNASTISFNOLDNMPDQATIGOKIOEPASKSANTVRGDDGLASAPETPTPTK 1167
 1837 IGIPHT-----ORLOGMPVKKHISIRKSAPLPSA----- 1867
 QY 1168 KESISSPAKLTSSAPKSPKIGISPVRYIKKNSIGIPIKATIKPKKKSQO--NEI 1225
 1868 ----AAHQOQLYGRSPSAVAMQAG-PRALAVQROMNMGVNLMPPAVNVSMNMNTLNM 1922
 QY 1226 SNHKVRDGGISPSGSEHQOHNPSMVSPSQY 1257
 1923 NSYRM-----TQPMNNSYHS-NNAVNMOTAOY 1949

RESULT 14
 DPQZ_HUMAN
 ID DPOZ_HUMAN STANDARD; PRT; 3130 AA.
 AC 060673; 043214;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA polymerase zeta catalytic subunit (Ec 2.7.7.7) (hREV3).
 GN REV3L OR POLZ OR REV3.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=98284025; PubMed=9618506;
 RA Gibbs P.E.M., McGregor W.G., Maher V.M., Nisson P., Lawrence C.W.,
 RT "A human homolog of the Saccharomyces cerevisiae REV3 gene, which
 encodes the catalytic subunit of DNA polymerase zeta."


```

QY 806 RRVSPMDLNLVSOFLPELSEDSGKDLNFANYSNNTN--RPRSTPLSTKNVLNIDN 863
DB 1082 HALISPPSPSYNAE-----FEDC---DLNTSDVMKSLGFLSERSTSPINS-----S 1124
QY 864 DPNVVEPPSPSYAEIRNARLSANKAPNOAPPLPPOROSPTSNSSNKKVRSRFRVPTF 923
DB 1125 PPRCSPPTDPR-----AEEIIMA--AAEKEAMLF--KGPVNYKKTVNSRIGKTSRARA 1172
QY 924 EIRRTSSALAPCDMYNDIFDDGAGSCKPTI--KAGMKTLRPMDDVYRIINAKGVNQ 981
DB 1173 QIKKSKAKLA-----NBSIVTKRKRKQTKLVDDGKKKPRAKO-KTN 1215
QY 982 DEYIATKLVDOPKRKNSTYTPDEDRYEELQTAISHNATIDS--YGRPDSISTDMLEY 1039
DB 1216 EKGTSRKHTTLKDEKIKSGAGAEVFKVKNVSEFASSGSGQLFKKD-----ML 1269
QY 1040 LSEDLKKP-----PVALSADRL-----FME-----QEVHP-LRSNSVLV 1073
DB 1270 MGSADVHPLASLPTGAINOQKLSGCFSSFLSEKSKSVDLQTPPSRDLHPVYVNCMI-- 1327
QY 1074 HPGAGAA-----TNSMLPEPDELINSPARNVSNNSDVVA-----ISGNASTISFN 1120
DB 1328 --GPEVSKINORPHNOSAMFTLKESTLQKNIFDLSNHLISOVAAQNTQISGSSKIEDN 1385
QY 1121 OLDMPDQATIGOKIOEQPASKSANTVAGDDGLASAPETPRTPTKESISSKP----- 1175
DB 1386 ANNTIRNTLSIG-KLSEVRNS-----LESKIDQAVTPPLFLCKXSOQOIVCI 1432
QY 1176 AKLSASPRKSPKIGSPVYVIRKNGSIAGI-EPTRKATHKPKKSGFQNEISNKKVRDGG 1234
DB 1433 AEOSSHSECSBPNATSEESQMPNCEVYLSRSPIKQIOMEOKORGFILDSMFKEPE-R 1490
QY 1235 ISPSGSEHQHNPAM-----VSVPQOYTA-----TSTVP 1265
DB 1491 VAPRSLSEISOTKALSOCKKRNSTPSPAFSGQSLAVKELLOKROQKAOANANTQDP 1550
QY 1266 DENKDVONKPRKOKKNNHNNH-----HHNKOKTDIPGVVDEDELDPVG 1310
DB 1551 LSNK---HQPKNKISGSLHNKANKRTRSVTSPRKRPTPRSRKOKKIKKL--KYDSJN 1605
QY 1311 LOERGL-----PFRVLGIR-----NINLPDIT-----HKGR 1338
DB 1606 LONSSQLNVSVDSPRIFPSDGFESCYSLEDSLSPENHYNP-DINTIGOTGFCSPYSGS 1664
QY 1339 FTULTNGVCHVTPPEYNNDDHNAIGKEFELTVADSLFEI-----LTKAS 1385
DB 1665 QVPAPADNL-----POKFLSD---AVODLFPQOALIEKNEFLSHDNOKCDEKHHTTDSAS 1716
QY 1386 YEKPRGTL-VEYTEKRVKVS--RNRLSRLEFGSKDITTTKFPTEVVDTWANKFAPDGSFA 1443
DB 1717 WIR-SGTLSPLEIFEKSTIDSNENR-----HNQKKNSHPLRTTS 1755
QY 1444 KCVTD---LOQFEDOTGKASQFDLNCFNEMETMSNGNOPMKRGKPYKTAOLEVKMLYVP 1500
DB 1756 NSIMSFVQOQAD-----CLSEKSKRLNRSYS----- 1783
QY 1501 RSDPRELLPTSRSAVESINELNNEQONFEGYLOHOGDCPIFKKRFKMGTSILAS 1560
DB 1784 -----KEYFLSLPQPN--SDWIOGHTKREKQSLDSANTSF---TALISSP 1825
QY 1561 EISHKTRAKINLSKYVDL-----IYVDKEN-----IDRSNHNFSVDLLDHAKIRF 1608
DB 1826 D-----GELYDVACEDELEYVSRNNMDLPTPDSPPRSTSSPSQSKKSGFTTPT 1874
QY 1609 ANGELIDFCAPKHKHEKIMI--QNLOEITYYRNRFRROP 1644
DB 1875 ANILKPLMSPSRREIMATLHDHSETIYOEPEFCSP 1912

```

```

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein HRX (All-1) (Fragment).
GN MLL OR HRX OR ALL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=C57BL/6J, and C57BL/6 X CBA; TISSUE=Spleen, and Lung;
RX MEDLINE=93317679; PubMed=8327517;
RA Ma O., Alder H., Nelson K.R., Chatterjee D., Gu Y., Nakamura T.,
RA Canani E., Croce C.M., Stracusa L.D., Buchberg A.M.,
RT "Analysis of the murine All-1 gene reveals conserved domains with
RT human All-1 and identifies a motif shared with DNA
RT methyltransferases."
RL Proc. Natl. Acad. Sci. U.S.A. 90:6350-6354(1993).
CC -! FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
CC MAY REGULATE GENES INVOLVED IN SKELETAL FORMATION DURING
CC EMBRYONIC DEVELOPMENT.
CC -! SUBCELLULAR LOCATION: Nuclear (by similarity).
CC -! ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -! SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -! SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -! SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -! SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
CC -! SIMILARITY: CONTAINS 1 CXC-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L17069; AAA62593.1; -
DR MGI: MGI:96995; M11.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003889; FYRICH_C.
DR InterPro: IPR003888; FYRICH_N.
DR InterPro: IPR003616; PostSET.
DR InterPro: IPR001214; SET.
DR InterPro: IPR002857; ZNF_CXHC.
DR InterPro: IPR001965; ZNF_PHD.
DR Pfam: PF00628; PHD; 3.
DR Pfam: PF00856; SET; 1.
DR Pfam: PF02008; zf-CXHC; 1.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00542; FYRICH; 1.
DR SMART: SM00541; FYRICH_N.
DR SMART: SM00249; PHD; 4.
DR SMART: SM00508; PostSET; 1.
DR SMART: SM00317; SET; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS50280; SET; 1.
DR PROSITE: PS01359; ZF_PHD_1; 3.
DR PROSITE: PS50016; ZF_PHD_2; 3.
DR DNA-binding: Nuclear protein; Zinc-finger; Metal-binding;
KW Transcription regulation; Alternative splicing; Polymorphism.
FT NON_TER 1
FT DNA_BIND 67 78 A.T HOOK (BY SIMILARITY).
FT DNA_BIND 115 125 A.T HOOK (BY SIMILARITY).
FT ZN_BIND 199 207 A.T HOOK (BY SIMILARITY).
FT ZN_FING 1044 1091 CXXC-TYPE.
FT ZN_FING 1330 1381 PHD-TYPE 1.
FT ZN_FING 1383 1432 PHD-TYPE 2.
FT ZN_FING 1465 1529 PHD-TYPE 3.
FT DOMAIN 1605 1650 BROMODOMAIN (DIVERGENT).
FT DOMAIN 3737 3846 SET.

```

```

FT DOMAIN 35 41 POLY-GLY.
FT DOMAIN 459 469 PRO-RICH.
FT DOMAIN 1231 1238 POLY-PRO.
FT DOMAIN 3533 3536 POLY-GLU.
FT DOMAIN 3693 3697 POLY-GLU.
FT VARSPIC 1503 1505 MISSING (IN ISOFORM 2).
FT VARIANT 1497 1497 K -> T.
SQ SEQUENCE 3866 AA: 420976 MW: ADFC55E14E806F1D CRC64;

Query Match 3.3%; Score 285.5; DB: 1; Length 3866;
Best Local Similarity 19.6%; Pred. No. 0.00087;
Matches 384; Conservative 249; Mismatches 756; Indels 575; Gaps 94;

QY 13 KHSLLQLOPQSSASIFSPKPLNFPRTNSKPSLDNPSSDYTSQDQK-KGKEEKD 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2095 RHSSLSPLRSKLRLM-SPV-----RTGSVSRSSVSPSLCTATNDPEASAKADRG 2147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   72 TAFQTSFDRNLDLNSIDIOQTIOHQ-----QQPQQQQLSOTD-----NLLDEFSEQ 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2148 GLSSSAMLGHASPPSSSSQRTVGSKTSHLDGSSPSEVKCSALDLVPKSLVGEKNR 2207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 TPMTSTLDLTQKON---PTVDKVENHAPTYINTSPNKSIMKATPKASPKKVAF---TV 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2208 TSSSKSTGSGASHAYPEIPRL---TPQVHNATPGLINIKISGFAPSPVPPSSKDTV 2263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 TNPETHYPRNRVEEDOSQOKEDSVF-----PLIOHQMKDPSQF 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2264 SYPLHL--LRGQSRDRDQMDPSSQVSKPSPNDEGIKTLKPGMGRPSILHEHIGSSSR 2321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 NYSEDTNASTVPPTPLTKTPTFAQ--LTKNNENYSEPEALTLQKLRKPNSTLDEK 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2332 DRQOKKSKSETCKEKKHSSKSTYLEPGQVTTGEGNLRPE-----FADEV 2366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 VN---LYLSPNNTNNNSKVNSDMSHLQNLQDASKNKTENIHLSFALKARKNDIENPLN 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2367 LTPGLGQRCPPNNVSSSEKIGD--KVLPLSGVPGQSTQ--VEGSKELQARKK----- 2415
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 332 SLTNADIS-LSSGSSSOSLOSLRNDNVLESVSPKPKV-----NPG--- 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2416 -CSVKYVPLMEGENOSK-----NPKESGSPSAHISVCPAEVVSASRSRPGAPG 2466
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 374 -----LSLNDGKIGFSEVVESSLPLDLSNDK----- 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2467 VQSPBNNTLSDPOSNNQNLPEODRNLMTIDGPKRPOEDGSFKRRTPRRSARAKNNMFG 2526
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 401 -----LETTKEHDAPEHNENFIDAKSTNTNKQQLVSSDHLDSFDRS---YNTHQ 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2527 LTPLYGVSRYGEEDIPFYSNT---GKKRGRKSAAGQVDGADLSTDEDDLYYNNPTR- 2582
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 451 SILMLNLSASQSLANLEKOROTQEOBOTQOAPREETSFSFNKVKQEPKSLFEVK 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2583 --TVISSGGERLASHNL-----FREEQCDLPKISOLDGVDDI----- 2619
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 511 VTIKEPVSAREIKA---PKR---EFSSRLIRIKNEDEIAPADI-----HPKKEENAN 558
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2620 -TESTSTYATISKRSSQYPKRNGKENGTEMLKIDRPEDAGEKENVYKSAVGHKNPEKLDN 2678
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 559 SHVEDTALLKALNDESDTONSTKMSIRFHIDSMDKLEDS-----NDGREDND 611
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2679 CH---SVSRVKAQGGDSLEAQLSSLESSRVTHTSPDKNLLDTYNALFLASDSDNNNSD 2735
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 612 DISREPKDIANDV---SQTSIDIQDKGNSSSEITTKLAPRPSDNNDKENSKLEDEPA 668
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2736 DCGNLPDIDMDFLKNTPPMQALGCESSSESLITLGEGLGLDSNREK-----DIG 2788
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 669 NNEISQOOLEVPHTKEDDSILANSNIPAPRELTPVVEANDYSFND----- 716
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2789 LFEVPSQOL--PATEPVDSV--SSSISAEQOFELPLPLPDLVLTTRSPVPSONSR 2844
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 717 -----VTKTFDAYSFEES-----LSREH-----ETDSKP 741
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2845 LAVISDSGEKRVTTTEKSVASSEGDPLALLSPGVDPAPRGHMTPDHFIQGHMDAHISSPP 2904
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 742 INFISIMWKOEKOKKH-----QIHKVPTKOILASVOQYKNEQESRYTSDKVKIPNAI- 793
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2905 CGSVEQHGNGNSODLTRNSGTPGLQVPSPVIVY-----QNGKYVBSSTDSPPSPQISMAAV 2960
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 794 -----QFKKFEVNMVMSRRVYSPMDDLNVQSFLE-----LSESGKRDNLFA 837
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2961 QTTPHILKPAATEKIL---VNNQNMOPLYVLOTLPNGVTKOILTSPVSSPVSMEYNTS 3016
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 838 -----NYSNNTNR--RSFTPLSTKNVLS-----NIDND 864
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3017 VLGRPGSGLTTLTGINPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 3076
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 865 PN-----VPEPERKSYAEIRNARR-LSANKAPQAPRLPQROPSPSTNSNKRVRSF 918
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3077 PSQLIGVQPPPPDQLLSEANQRTDLTTVATPSSG---LKKRPISRLHFRKKKLAPS 3133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 919 RVPPFEIRTSALAPCMYNDI-----PDDGAGASKPIIKKGM-----KLPSMKDD 968
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3134 SAP-----SNIAPSVSNMTLINFPPOLSNNPRLDLGLSLNPSGHRTVPNI--- 3181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 969 VKRLNARKGYTODEYINAKLVQDKPKKNSIVTDPEDRYEELQQTASIHNTIIDSITYGR 1028
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3182 IKR---SKSGIMYFE-----QAP-----LLRP-----QSVGTATTAAGS----- 3213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1029 PDSISTMLPYLSDELKPPALLSADRLF---MEQVHPLRSNSVYVHGAGATSS 1084
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3214 -STISOD---TSHLTSGPVASLAGSSVVLNVWSQTTAAPTSTSY---PGHYTLANOR 3265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1085 MLPEPDFE---LINSFARVNSNNDVVALSGNASTISFQOLMNFDDQ---TIGOKI 1136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3266 LIGTPDISTSILHLKASHQSLGIDQDVALP--PSSGMFQOLGTSQTPSAAMTAASI 3323
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1137 QEOPASKSANTVRGBDDGLA-----SAP----- 1159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3324 CVLPSSQTAAGMTAASPGEAEENHYKLGQNLQLAGKTGLTSQRDRDPSAPGQPSIFT 3383
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1160 ETPRTPT-----KESISSRPAKLSSAPKSPKSPKSPKSPKSPKSPKSPKSPKSPKSP 1213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3384 QTAEPNGVSLQONTLPUSAKPA--SSASPGSSP-----SSQOQSSSVSPPT 3430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1214 H-KPKKS-----FQNETSNKVRDGGISPSGSHQOHNPSVSPQOYDATSTVDE 1267
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3431 KPKKAKRIOLPLDKGSKKKKHYSHLRTSSSEAHNLRHRTDPA-----PQSPYTRIPRARE 3486
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1268 NKDYOKKPREKOKO-----KHNHNNHNNHNNKOTDIPGVVDEIDVGLQERK 1316
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3487 QODAGVQOPQKECGQAGVUALREVQATONPRANQENAEPRKAMEBE--ESGFSSPLM 3544
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1317 LFFRYLGKINLNPDIINHGR--FTLLIDNGVHCYTPPEYMMDHNAIGKFEFLVADS 1375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3545 LMLQOEOKRKESTIERKPKGLVFEISSDG-----FQI-CAES 3582
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1376 LEFI---LTLKASYEKPRTGLVEVEKRYVKSRN-----RLSLRFGSKDITTTK 1422
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3583 IEDAKMSLTDVQOERSNARKOLKOLSFAGVNGLRMIGLIDAVVELIDLAQKH-CRNYK 3641
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1423 F---VPTVEKDTMANKFAPDGSFARCYTIDQOFEDQITGKASQEDLNGFNCMEWETMSNGQ 1479
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3642 FRFHKPEEANBPPLN---PHGS-ARAEVHLRK-----SARDMEFN-----LASKHRQ 3684
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1480 PMKRGKPKYKIQOLEVKMLYVRSRDPRELTPTSISATYSIMELNNEQNNYFEGYLHDEGG 1539
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3685 PPEY--NPNDDEEEVEQLSARATSM-D-LPMPMR--FPHLKRTSKSAVGVRSPHGG- 3739
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1540 DCPTEKRFKFLMGSLLAHS-----ELSHKTRAKINKSKVVDLYV---DKENIDRSNH 1591
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3740 ---LECKRNID-AGEWVLEYGNVIRSIQTDKREYVDSKIGCYMERIDISEVVDATMH 3795
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1592 RNFSVLLLDHAFAKIKFANGELIDFCAPNKHENKIMIONLOEII 1635
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3796 GNA-----RFINHS-----CEPNCTSRVYINIDGQKHV 3824
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Mon Mar 17 12:50:24 2003

us-09-964-858-1.rsp

Page 21

Search completed: March 17, 2003, 12:24:53
Job time : 105.724 secs

This Page Blank (uspto)

STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/264, 604
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/642, 846
FILING DATE: 03-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: MUETING, ANN M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110, 00280101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ. ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1664 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-264-604-2

Query Match 100.0%; Score 8631; DB 4; Length 1664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNSTPRLPIDKSHLQLOPOSSASIFSPKPLNFPTNKKPSLIDPSSSDTYTSEQ 60
DB 1 MNSTPRLPIDKSHLQLOPOSSASIFSPKPLNFPTNKKPSLIDPSSSDTYTSEQ 60
QY 61 DQKGEKKEDTAFQTSFDFNFDLNSIDTQITQHQOQPOQOQOQLOQDNNLIDEF 120
DB 61 DQKGEKKEDTAFQTSFDFNFDLNSIDTQITQHQOQPOQOQOQLOQDNNLIDEF 120
QY 121 QTPMTSTLDTKONPTVDKYNENHAPTYINTSPNKSTIMKATPKASPKKVAFTV 180
DB 121 QTPMTSTLDTKONPTVDKYNENHAPTYINTSPNKSTIMKATPKASPKKVAFTV 180
QY 181 HYDNNVEEEDSOQKEDSEVERPLIOHOKDPSQVNSDEDTNAAVPTPTTTPK 240
DB 181 HYDNNVEEEDSOQKEDSEVERPLIOHOKDPSQVNSDEDTNAAVPTPTTTPK 240
QY 241 QLLNNKNEVSEPEALTDMLKRENFSLSLDEKVNLYLPTNNNSKNVSDMSHL 300
DB 241 QLLNNKNEVSEPEALTDMLKRENFSLSLDEKVNLYLPTNNNSKNVSDMSHL 300
QY 301 QDASKNKTENIHNLSFALAKPKNDIENPLNSLTNADISLRSSGSSOSSLO 360
DB 301 QDASKNKTENIHNLSFALAKPKNDIENPLNSLTNADISLRSSGSSOSSLO 360
QY 361 ESVPSPKPVNPGSLNDGKGSDEVESLLPRDLSRDLKLETTKEHADPEHNEN 420
DB 361 ESVPSPKPVNPGSLNDGKGSDEVESLLPRDLSRDLKLETTKEHADPEHNEN 420
QY 421 KSTNTKKGOLVSSDDHLSFDRSYNHTEOSILNLNSASOSQISLNALEKOR 480
DB 421 KSTNTKKGOLVSSDDHLSFDRSYNHTEOSILNLNSASOSQISLNALEKOR 480
QY 481 TQAAPEEETSFSDNITKVOEPSNLEFVKVTIKKEPVSAATEIKAKRRE 540
DB 481 TQAAPEEETSFSDNITKVOEPSNLEFVKVTIKKEPVSAATEIKAKRRE 540

QY 541 DEIAEPADIHPPKKNENANSHVEDTDALLKALNDEESDTQNSTKMSIRFH 600
DB 541 DEIAEPADIHPPKKNENANSHVEDTDALLKALNDEESDTQNSTKMSIRFH 600
QY 601 DSDNGREDNDOLSRFKSDILNDVQSDITIGDKGNSSSETTTTTLAPPSD 660
DB 601 DSDNGREDNDOLSRFKSDILNDVQSDITIGDKGNSSSETTTTTLAPPSD 660
QY 661 SKSLPEPPANNESIQOOLEVPHTEKSDSILANSNINAPPELTLPVEAND 720
DB 661 SKSLPEPPANNESIQOOLEVPHTEKSDSILANSNINAPPELTLPVEAND 720
QY 721 FDAYSFEESLSREHETDSPINFISIMHKOEKKHQIHKVPYKQIIAS 780
DB 721 FDAYSFEESLSREHETDSPINFISIMHKOEKKHQIHKVPYKQIIAS 780
QY 781 RVTSDDVKIIPNATQFKKFEVNVMSRRVSPMDDLNVSOFLPELSEDS 840
DB 781 RVTSDDVKIIPNATQFKKFEVNVMSRRVSPMDDLNVSOFLPELSEDS 840
QY 841 NNTNRPSFTPLSTKKNVLSINIDNPVVEPEPKSYAEIRNARLSANKA 900
DB 841 NNTNRPSFTPLSTKKNVLSINIDNPVVEPEPKSYAEIRNARLSANKA 900
QY 901 QROPSTRSNKNKRVSRFRVPTFEIRTSALAPCDMYNIDEDGAGAK 960
DB 901 QROPSTRSNKNKRVSRFRVPTFEIRTSALAPCDMYNIDEDGAGAK 960
QY 961 LPSMDDDVKIRILNAKKGYTODEYINAKLVYDQPKKNSIYTPDE 1020
DB 961 LPSMDDDVKIRILNAKKGYTODEYINAKLVYDQPKKNSIYTPDE 1020
QY 1021 IDSSIVGRPDSISTDMLPYLSDELKRPALLSADRLFMEQEVNPLRS 1080
DB 1021 IDSSIVGRPDSISTDMLPYLSDELKRPALLSADRLFMEQEVNPLRS 1080
QY 1081 TNSMTPEDDFELINPARVNSNNSDVAISGNASTISFNQDMDADAT 1140
DB 1081 TNSMTPEDDFELINPARVNSNNSDVAISGNASTISFNQDMDADAT 1140
QY 1141 ASKSANTVGRGDDGLASAPETPTPKKESISSKPKAKLSASAPRKPI 1200
DB 1141 ASKSANTVGRGDDGLASAPETPTPKKESISSKPKAKLSASAPRKPI 1200
QY 1201 GSTAGIEPIPKATHKPKKSFQGNESINHKVRODGLSPSSGSEHQ 1260
DB 1201 GSTAGIEPIPKATHKPKKSFQGNESINHKVRODGLSPSSGSEHQ 1260
QY 1261 TSTVPRDKNDVQHKPREKQKKNHNNHNNHNNHNNHNNHNNHNNH 1320
DB 1261 TSTVPRDKNDVQHKPREKQKKNHNNHNNHNNHNNHNNHNNHNNH 1320
QY 1321 VLGIRKINLPDINTHNGRTLTLDNGVNCVTTREYMMDDHNAIGKE 1380
DB 1321 VLGIRKINLPDINTHNGRTLTLDNGVNCVTTREYMMDDHNAIGKE 1380
QY 1381 TLKASYEKPRGTLVEYTEKRVKSRNRLSRLFGSKDITTTKEVP 1440
DB 1381 TLKASYEKPRGTLVEYTEKRVKSRNRLSRLFGSKDITTTKEVP 1440
QY 1441 SFARCTYIDQOEFDDITGKASQFDLNCFMWETMSNGNPMKRG 1500
DB 1441 SFARCTYIDQOEFDDITGKASQFDLNCFMWETMSNGNPMKRG 1500
QY 1501 RSDPREILPTSTRSAVESINELNEDNNYFEGYLDGEGDCITFK 1560
DB 1501 RSDPREILPTSTRSAVESINELNEDNNYFEGYLDGEGDCITFK 1560
QY 1561 ELSHKTAKINLSKVVLDIYVDKENIDRSNHRNFSVLLLDIAFAK 1620
DB 1561 ELSHKTAKINLSKVVLDIYVDKENIDRSNHRNFSVLLLDIAFAK 1620
QY 1621 KHEMKIWNLOEITTYRNRFRROPVNVNMLQOQOQOQOQOQOQO 1664

Db 1621 KHEMKWIONLOETLYRNRFRROPVNLML000000000SSQ 1664

RESULT 4
US-09-599-652-3

; Sequence 3, Application US/09599652
; Patent No. RE37741

GENERAL INFORMATION:

APPLICANT: HOSTETTER, MARGARET K.

APPLICANT: GALE, CHERYL A.

APPLICANT: BENDEL, CATHERINE M.

APPLICANT: TAO, NIAN-JUN

APPLICANT: KENDRICK, KATHLEEN

TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE

TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MUETING, RAASCH, GEBHARDT & SCHWAPPACH, P. A.

STREET: 119 NORTH FOURTH STREET, SUITE 203

CITY: MINNEAPOLIS

STATE: MINNESOTA

COUNTRY: USA

ZIP: 55401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/599,652

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/642,846

FILING DATE: 03-MAY-1996

ATTORNEY/AGENT INFORMATION:

NAME: MUETING, ANN M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 110.00280101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1228

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 236 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FEATURE:

OTHER INFORMATION: amino acid positions 218-453 from SEQ ID NO:2

US-09-599-652-3

Query Match 14.0%; Score 1211; DB 1; Length 236;
Best Local Similarity 100.0%; Pred. No. 5.5e-67;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 218 SDEDINASVPPPTPLHTTKPTFAQLLNKNEVNSEPEALTDMLKRENFNSLSLDEKYNL 277

Db 1 SDEDINASVPPPTPLHTTKPTFAQLLNKNEVNSEPEALTDMLKRENFNSLSLDEKYNL 60

Db 278 YLSPTNNNSKNVSDMDSHLQNLQDASKKNTNENHNLSEFALKAPKNDIENPLNSLTNAD 337

Db 61 YLSPTNNNSKNVSDMDSHLQNLQDASKKNTNENHNLSEFALKAPKNDIENPLNSLTNAD 120

Db 338 ISLRSSGSSQSLQSLRNDNRVLESVPGSPKKVNPGLSLNDGIKGFSDEVESLPRDLS 397

Db 121 ISLRSSGSSQSLQSLRNDNRVLESVPGSPKKVNPGLSLNDGIKGFSDEVESLPRDLS 180

Db 398 RDKLETTKEHDAPEHNNEFIDAKSTNTNKGQLLVSSDHLDSFDRSYNHTEQSIL 453

Db 181 RDKLETTKEHDAPEHNNEFIDAKSTNTNKGQLLVSSDHLDSFDRSYNHTEQSIL 236

RESULT 5
US-08-642-846-3

; Sequence 3, Application US/08642846
; Patent No. 5886151

GENERAL INFORMATION:

APPLICANT: HOSTETTER, MARGARET K.

APPLICANT: GALE, CHERYL A.

APPLICANT: BENDEL, CATHERINE M.

APPLICANT: TAO, NIAN-JUN

APPLICANT: KENDRICK, KATHLEEN

TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE

TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MUETING, RAASCH, GEBHARDT & SCHWAPPACH, P. A.

STREET: 119 NORTH FOURTH STREET, SUITE 203

CITY: MINNEAPOLIS

STATE: MINNESOTA

COUNTRY: USA

ZIP: 55401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/642,846

FILING DATE: 03-MAY-1996

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: MUETING, ANN M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 110.00280101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1228

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 236 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FEATURE:

OTHER INFORMATION: amino acid positions 218-453 from SEQ ID NO:2

US-08-642-846-3

Query Match 14.0%; Score 1211; DB 2; Length 236;
Best Local Similarity 100.0%; Pred. No. 5.5e-67;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 218 SDEDINASVPPPTPLHTTKPTFAQLLNKNEVNSEPEALTDMLKRENFNSLSLDEKYNL 277

Db 1 SDEDINASVPPPTPLHTTKPTFAQLLNKNEVNSEPEALTDMLKRENFNSLSLDEKYNL 60

Db 278 YLSPTNNNSKNVSDMDSHLQNLQDASKKNTNENHNLSEFALKAPKNDIENPLNSLTNAD 337

Db 61 YLSPTNNNSKNVSDMDSHLQNLQDASKKNTNENHNLSEFALKAPKNDIENPLNSLTNAD 120

Db 338 ISLRSSGSSQSLQSLRNDNRVLESVPGSPKKVNPGLSLNDGIKGFSDEVESLPRDLS 397

Db 121 ISLRSSGSSQSLQSLRNDNRVLESVPGSPKKVNPGLSLNDGIKGFSDEVESLPRDLS 180

Db 398 RDKLETTKEHDAPEHNNEFIDAKSTNTNKGQLLVSSDHLDSFDRSYNHTEQSIL 453

Db 181 RDKLETTKEHDAPEHNNEFIDAKSTNTNKGQLLVSSDHLDSFDRSYNHTEQSIL 236

RESULT 6

US-09-264-604-3

; Sequence 3, Application US/09264604

Patent No. 6346411
GENERAL INFORMATION:
APPLICANT: HOSTETTER, MARGARET K.
APPLICANT: GALE, CHERYL A.
APPLICANT: BENDEL, CATHERINE M.
APPLICANT: TAO, NIAN-JUN
APPLICANT: KENDRICK, KATHLEEN
TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: MUETING, RAASCH, GERHARDT & SCHNAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/264,604
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/642,846
FILING DATE: 03-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: MUETING, ANN M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110,00280101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 236 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
OTHER INFORMATION: amino acid positions 218-453 from SEQ ID NO:2
09-264-604-3
Query Match 14.0%; Score 1211; DB 4; Length 236;
Best Local Similarity 100.0%; Pred. No. 5,5e-67;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SDEDTNNAVPPRPPLHTTKPTFAQLLNKNEVNSEPEALTDKMLKRENFSLDEKYNL 277
DB 1 SDEDTNNAVPPRPPLHTTKPTFAQLLNKNEVNSEPEALTDKMLKRENFSLDEKYNL 60
QY 278 YLSPNNNNKRVSDMDSHLQNLQDASKKKTNTNHNLSFALKAPKNDIENPLNSLTNAD 337
DB 61 YLSPNNNNKRVSDMDSHLQNLQDASKKKTNTNHNLSFALKAPKNDIENPLNSLTNAD 120
QY 338 ISLRSSGSSSLQSLRNDNRVLESVPGSPKKVNPGLSLNDGKSGDEVEVSSLLPRDLS 397
DB 121 ISLRSSGSSSLQSLRNDNRVLESVPGSPKKVNPGLSLNDGKSGDEVEVSSLLPRDLS 180
QY 388 RDKLETTKEHDAPEHNNENFIDAKSTNTNKGQLLVSSDDHDSFDRSYNHTEOSIL 453
DB 181 RDKLETTKEHDAPEHNNENFIDAKSTNTNKGQLLVSSDDHDSFDRSYNHTEOSIL 236

RESULT 7
US-09-134-001C-3159
Sequence 3159, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:

APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NOLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3159
LENGTH: 10182
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 4.3%; Score 375; DB 4; Length 10182;
Best Local Similarity 17.9%; Pred. No. 1.5e-13;
Matches 363; Conservative 326; Mismatches 688; Indels 650; Gaps 91;

QY 34 KPLNFPRTSKPSLDPNSSD--TYSBODKKGKREKDD-----TAFQTSFDRFELD 85
DB 3424 EPLETAKNLOANIDOKPSTDGMTOOSVYERKIQEAKDKINSTNNVLANPVDYNAIRT 3483
QY 86 NSIDIQOTIQHQOQPQO-----QQQLSQTNNLIDFSFQTPMTSTDLTKQN----- 134
DB 3484 NKVEIEQ-INNELTQAKQGLTVDKQPLINAKTALQSLDNOGSTGTMETATONYNAKRQ 3542
QY 135 -----PTVDKYNENHAPRYINTSPRKSTIMKATKASKKKKAFVYTNTEIHNYDNRE 188
DB 3543 KAEQYIQANAKKITEENAPSVQGVSDSKVEQALSELNNAKSALRADKQEOLOQAVNYQLD 3602
QY 189 EEDSOQKEDSVPEPLIQHWKMDPSQFNYSDEDTNNAVPPRPPLHTTKPTFAQLLNKNE 248
DB 3603 PTDLNKKKRASTITAYNRQ-----QFSNELNSTKTITDIIIEQNSVADVNNALNK 3655
QY 249 VNSEPEALTD---MKLRENFSLDEKYNLYLS--PTNNNSKRVSD----- 292
DB 3656 VREVOQKLEARALLQNKEDNSALVRAKEQLOQAVDQVSTEGMTQGFQDYNSKQAAO 3715
QY 293 -----MDSHLQNLQDASKKKTN-----ENIHNLSFALKAPKNDIENPLNSLT-NAD 337
DB 3716 QEISKAQOVIDNGDATVTDQISNAKTNERALPALNNKATGLADKDEELQNAVNYQLDND 3775
QY 338 -----ISLRSSGSSSLQSLRNDNRVLESVPGSPKKVNPGLSLNDG 379
DB 3776 TSGKTPASIRKRYNEAKSRQIQIDSAKKNANSILTNNDNQSVYTAALNKI----- 3826
QY 380 IGFSDVEVESLLPRDLSRDKLETTKEHDAPEHNNENFIDAK----- 421
DB 3827 -----KAVQPELDRAIAMLK---NKE-----NNNALVQAQOQLOQIYNEVDPTQGMTT 3871
QY 422 STNTNKGQLLVSSDDHDSFDRSYNHTEOSILNLSNASQOSLSINALEKOROTEOQ 478
DB 3872 DTANNYKSKKRAEDELQKAOQIINNCGATFEOQITNETNRVMO---ATINAIKKANDLRA 3928
QY 479 EOTOAEPREETSFSDNKIKVKQEPKSNLEFVYVTKKEPVSAETIKAPKREKSSRIIRIK 538
DB 3929 DKSQI-----ENAYNQILQ-----NVD-----TNGKKPASTIOQVQAARQALETOYNNAK 3972
QY 539 NE-----DEIAEPAD-----IHPKKENANSHVEDTDALEKKALND 574
DB 3973 SEAHQITLNSNPSVNEVQAOLQKVEAVOLKVDALHIIHQNKENSALVATAKNOLOOSVND 4032
QY 575 D-----EESDTQNSTKMSIRPHIDSKMLDSNODREDNDISFEKS-DIL 622
DB 4033 QPLTGMTODSINNYEAKRNEAQSAIR---NNAEAVINNGDATKQISDEKSVVEQALHL 4089
QY 623 NDVSO--TSDIIG-----DKYNSSESSEITTKTLAPPSDNNKREKNSLE----- 665
DB 4090 NDAKOQITADTTELOTFAVOQLNRRDGTNNK-----KPRSTINAYNKAIOSLQSLFQITSAK 4142

QY	666	DPAN-----	-----	NE5LQO-Q	LEVPHTEKEDSILANSN	TAPEBLLPVE	AND	710																																														
Db	4143	DNAAVIQA	PIRIVQEV	ENALLOO	NOLO	NOLO	TAI	NOLOPLS	NDA-----	LKAAR	4196																																											
QY	711	YSSFNDDYKT	-----	FDAYSSFEES	LSREH-----	-----	ETD-----	738																																														
Db	4197	ENKINQV	OVQODMQOSIE	AVQNAK	RAQON	ESMTALAL	INNGDAD	QUTTET	EDRVNOQT	4250																																												
QY	739	---SKP	INFSIMHKQ--	EKOK--	KHO	IKHP-----	TKOI	IASYOQ--	YKNE---	777																																												
Db	4257	TNLTOA	INGILTV--	NKEP	LETA	KTALONN	IDOVST	FGMTQOS	AVANYOK	LOAKMEINTI	4315																																											
QY	778	-----	QSR	TSQK	VKIP	PA	IAQ-----	EKKFK-----	799																																													
Db	4316	NNVLANN	DVNAIK	TNAE	ERIS	NDLTQAK	NNLO	QDTP	LEKIK	ROLODE	IDOGT	MTDG	43757																																									
QY	800	-----	EVN	MSRRV	SP	DMDL-----	NVSQ	FPEL-----	825																																													
Db	4376	MTQOSV	ONYNDLS	LAILE	EKG	KNKLLK--	MTVQ	QV	ESVANA	QOVLQ	QNA	R	TSIV	44330																																								
QY	826	-----	SED	S	GR	FDL	MFANS	NTNR	PRS	FTPL	STR	NVLS	NI	NDP	865																																							
Db	4434	PDKTQ	Q	EA	K	R	LENS	INOQ	TD	FG	MTQOS	L	N	Y	ND	K	LAKARQ-----	NLEK	IS	VL	GQ	44889																																
Y	866	NVPE	PERPK	SAE	ER--	NAR	L	S	A	N	K	A	P	NOA-----	P	E	L	P	PO	Q	S	T	S	N	K	R	V	R	P	P	E	923																						
Db	4490	TV-----	AEI	R	O	N	T	D	E	A	N	A	K	A	L	D	A	R	S	O	L	N	R	E	P	I	N	H	I	N	S	H	N	A	K	O	N	F	45400															
QY	924	EIR	T	S	A	L	A	P-----	CD	W	N	D	I	P	D	E	G	A	S	K	P	T	I	A	E	G	K	T	P	S--	MD	D	O	V	K	R	971																	
Db	4541	KAOV	N	S---	AP	N	H	T	L	E	T	I	N	K	A	D	T	L	N--	Q	S	M	T	A	L	S	E	I	A	D	E	N	O	K	O	E	N	T	45904															
QY	972	-----	IL	N	A	K	G	V---TODE	Y	I	N	A	K	L	V	D	O	KP--	K	N	S	I	V	T	D	E	D	R	E	E	L	O	T	A	S	I	N	A	T	10200														
Db	4595	QDY	N	A	N	A	N	A	K	G	I	L	N	O	T	S	P	T	M	S	A	D	V	I	D	O	K	A	E	D	V	K	R	T	A	L	D	O	N	Q	L	E	V	A	K	O	A	L	N	H	L	N	T	46504
QY	1021	ID	S	I	G	R	P	O	S	I	S	T	D	M	L	P	Y	L	S	D	E	K	R	P	P	A	L	S	A	D	L	E	M	E	O	V	N	H	R	L	S	V	L	H	E	G	A	10808						
Db	4655	LND-----	-----	LND	A	O	R	O	T	L	D	T	I	N--	-----	H	S	P	N	I	S	V	A	O	A	E	K	A	N	T	46888																							
QY	1081	TNS	G	M-----	L	P	E	P	E	L-----	I	N	S--	P	A	R	V	S	N	S	D	N	A	L	S	G	A	S	T	I	S	P	O	D	M	N	D	11277																
Db	4689	VNT	A	M	T	O	L	K	O	T	I	A	N	Y	D	E	L	D	E	L	D	G	V	I	N	A	N	D	K	A	Y	A	N	A	N	A	N--	N	A	K	L--	I	N	O	S	D	A	N	O	47420				
QY	1128	--DO	A	T	G	O	K	I	O	E	O	P	A	S	K	A	N	T	V	R	G	D-----	D	G	L	A	S	A	P	E	T	P	R	T	P	T	K	E	S	11772														
Db	4743	Q	L	D	A	E	I	N	K	T	O	V	A	N	T	K--	N	D	L	N	G	D	K	L	A	E	A	K	R	D	A	N	T	I	D	G	L	T	Y	L	N	E	A	O	R	U	K	A--	K	E	N	G	47990	
QY	1173	S	K	P	A	L	S	S	A	P	R	K	S	P	I	G																																						

```

Db 5120 TLDASINQKAAILTTTNALDGEQLRRK-----ENADEI----- 5157
QY 1515 AYESINEINNEONNFFEEGLHOEGDCPIFK-----KRFKLMGTSLAHSEISHKTRAK 1569
Db 5158 --NTLNQJLTDQARNSEKELINSSQRTREYASQALAKELNKVM-----EQNLHNLGK 5208
QY 1570 ---INLSKVVDLIYDKENIDRSNHRNFSVDYLLDHAERIKFANGEL 1613
Db 5209 NQMINSSKFI-----NEDANQQAAYSNAIASAEALKNKSQNPQL 5247

RESULT 8
US-09-134-001C-5080
: Sequence 5080, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NOCIDEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134, 001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 5080
: LENGTH: 3696
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 4.0%; Score 348.5; DB 4; Length 3696;
Best Local Similarity 18.2%; Pred. No. 1.6e-12;
Matches 363; Conservative 327; Mismatches 759; Indels 549; Gaps 90;

QY 1 MNSPFSKLLPDKHSHLOLOPQSSASIFNSPTKPLNPPRPNRSKPSLDPNSSSDPTTSEQ 60
Db 1187 LNSPTGHAQD-EKQDALIRLQAKETAL-----NDINQAQINQVNDTALTGIGNIQNTQ 1240
QY 61 DOEKKEKKQDTAFQTSFDRNFDLDSIDIOQTIOHQOQRPQOQOQSLQTDNNLIDFSF 120
Db 1241 VNVRRKQQAQKTTI-----NDI-VQOHKOSIQNNDDATEEKEVANNILVN--AS 1285
QY 121 QTPMSTSLDLKQNTVQKVENHNAFYINSPKNSIKKATP-----KASKKVAFTYTN 176
Db 1286 QONVTSKIDNATTTNQIDIGVISDQROSNATTPRSTIRKKNKDIDIKAAQKKI----- 1339
QY 177 PEIHNYPRNRYEEDOSQOKEDSVEPPLIQHQMKPDSQFNYSDEDTNAS-----VPPT 229
Db 1340 -KIQGINATDEELQEAHRK---IEAKIEA-KDNIDRNSTRQVYNKAKINGINKIENTI 1393
QY 230 PRLATTKPTFAQLL-NKNN-----VNSEPEALTDK----- 260
Db 1394 TPATTVKSEARQAVQNKANEQINHTQNTPRDATNEEKQRAINRVSASELARVOAQINAEHTT 1453
QY 261 -----LKEENFSNLS-----LDEKYNLYLSTPTNNNSKNVSDMOSH-L-QN 299
Db 1454 QGVKTIKQDADTSLSRINAQVVEKESARNAIEOKATQTOQTOPIINND--NATDEEEVANN 1511
QY 300 LODASKKNTNENINHLSPALKAPKNDIENP-----LNSLTNNDISLRSSGSSQSSLSQSLRN 355
Db 1512 LVIAIRKQKSLDNINSL-----SNNDVENAKVAGINEIAN----- 1546
QY 356 DNRVLESVGSPPKVNPSLSINDIGKFSDEVESLSLRDLRSRKLLETKEHDAP----- 410
Db 1547 -----VLPAATVAKSKAKKDIDQKLAQOI-NQIQTHQATTEKEKAIAIOLAN 1591
QY 411 -----EHNNEFIDAKSTNTNKKGOLLV-----SSDHLDSFDRSYNHTQSLTN 454
Db 1592 QKSNEMARPAIQNEHNSNNGVAAQKNSGNGIHEIRLVNMDAKKSDAKQSIDNKTN--EQS--N 1647

```


QY	455	LLA-----SOSISNALE-----KQRTQOQOBTQAAPEE-ETSSDIKRYQ	500
Db	1648	TINTPDATBEOKALDKLAKDAGYKNVDAQNQQOVSAKTEADITTTINOANAK	1707
QY	501	EPKSNEEF-KVAVIKKEPEVSATEIKAPKEEFSRILIRINEDEIAPADIPHKRENEANS	559
Db	1708	KPSARVELDSKFDLKNQIWAATP-NMTEEKQDAIQRLOGKRDEYKNLINDORRNEVEQ	1766
QY	560	H----VEDTALLKALINDEESDTTQNSTKSIIRPHIDSMKLEDSNDGREDNDISR	615
Db	1767	HNIGIOLEETI---HANPFRKSDAQO---ELQTF-ISOTELINNNDATNEEKDEAKR	1819
QY	616	---FEKSDIINDY--SOTSIDIQDKGNSSSEITP-----KTL	648
Db	1820	LLEISKKTTITNINOQTNOQNMADNMGNEIATITPATTIKTAKRALDKKAQOQYTI	1879
QY	649	APPBSNDNKENS-----LEDPANNESIQQOLEVPHTEKEDDSILANSSNITAPEE	700
Db	1880	INGNNDATDEKEAKRLYEKAKIEKSNITNSDPEREVNGAKTNG--LEKINNIPSTQ	1937
QY	701	LTLPVNANDYSSPND-----YAKTFDVASSF-----ESLSREHT	737
Db	1938	---TTNNAKOEINDKAQOLIOINTPDPAIEEKOETNRVNAGLAOIIONINNAHST	1992
QY	738	D-----SKPINFISIMHKOKO-----KHQIHVP	763
Db	1993	QEVNESKTSNIATIKSVQPVNIKKPAINSLTQOEAQNNQKTLIGNGNATDEKEAKOLV	2052
QY	764	TKOIIASVQO-YKNEDESRYSDK-----VKIPNALQKKKKEVNVASRRVSPMDML	816
Db	2053	TOKNEOIDQIHSTQONOVQDNVKAQAITALKILINNAHKRODAINILTL-NALESKSDI	2111
QY	817	NVSOFLPELSEDSGFKDLN--FANYGNTNTRRPSFPLSTKVLNLSINDPVRVPERPK	874
Db	2112	RANQDATTEKKNATIQSIDTLQANNNINGANT-NALVDLEBGOKKQKQIVLESTQK	2170
QY	875	SVATIRNARLRSANKAPRNOAPPLPQORQSSRFSRNSKNRVSRFRVPPFEIR-KTSSALA	933
Db	2171	TOAKADIAQIIGORSTIDP-----NOMATTEKQALRLQOENGVNDRIQOALA	2222
QY	934	---PCDYNIDPFDFGAGSKPTI-----KAEGMKPLPSMDKO--DVRILINA	975
Db	2223	NQNTDEKKNMILETI-RNVEPIYIVKRAHEIKRKAAGOTLLINONODATILEEQIALG	2281
QY	976	KKGYTODEYIN-----AKLVQDKP-----KNSIYTPDERYEE	1009
Db	2282	KLEEVKMEALNOVSOAHSNNDVKAIEANGIAKISEVHEPETIIRKNAKOEIEDOASQIDT	2341
QY	1010	L---OQTASIHNAITDSSITYGRPSIS-----TDMRLPYLSE	1043
Db	2342	INANKSTNEKSAIDRVAVAKIDAINNITNATTTQLVNAKNGSNTSISQILP--STA	2399
QY	1044	LKKRPYALLSADRLFEMOEYVHLRSLSVLH-PGACA-----ATN--SMLPEPDPFLIN	1095
Db	2400	VKTALALASE-----AKNKWAIIDQRPNATAEKEBEANKVDRLOEADANILK	2450
QY	1096	SPANVSNNSDNVAISG-NASTISF-----NODMNPDDATIGAKIOQEPASKA-	1145
Db	2451	AHTTDEVNNIKNAVOYNINAVUEVIAKONVKKNOLOFIDNKKITENTPRATTEEKEA	2510
QY	1146	---NVPRDDGLASAPETPTPTPKKESISKRAKLSASAPKSPKIGISPVUKIN	1200
Db	2511	NRLLONLVLTSTSEIANDVHN-----NEVDALDKAR-----KIEIYOVASK	2555
QY	1201	GSJAGIEPRKATNPKKSFQGNELSNHKVURDQGISPSSGSEHQDNPMSVPSQUTDA	1260
Db	2556	RDV--LVAIOEAFNSQTOEIOENQOATNEKEFALN---KINOLLQAKVINDO----	2604
QY	1261	TSVPRDNKKV-----QNRPREKQOKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	1305
Db	2605	---AQSNCVDSAKTRSDIODIEOIORPHQOTATGNGHNRKNEKANOQOSTATNPRSTEE	2660

```

OY      1306   IPDVG--LOE-----RGK-----LFFRYLGIKNLP-----DINH 1335
                |||               :||
Db       2661  ROEASAKIQOEVLKKAIKIDKQTDNDDEKIYVNGIAEINI-LPRTTYKKAKADAVNAE 2719
                |||               :||
OY      1336  KGRFTLLTLDNGVCYTTPREYNMDHNVAIGKEFEELTVADSLFELL-PLKASYERPRGTL 1393
                |||               :||
Db       2720  KEOKMLQINSNDEAT-----BEKLIVASDNLNHVETTTNOALIEDAPDTNQ 2764
                |||               :||
OY      1394  VEVTERRK-----VKSRLNLSRFLGS-----KDITTFKFVPLEVVDYANANKA 1437
                |||               :||
Db       2765  VNVEKNKGIGTRIDIOPLVKKPPTAKSKIESAVEKKTEINOTOMATHDEVRE-GLINOLN 2823
                |||               :||
OY      1438  PDGSEARCYIIDLOEQEDDITKASOFDINCEN-----EMETWSNGNOPMKRCGPYRIAOLE 1493
                |||               :||
Db       2824  QJHEKAKNDVNQSQNQOVE-NAEQNSLSDQINNFRPDFSKRNNAVAYELIQAQONNIDELE 2882
                |||               :||
OY      1494  VKM--LYPRSPRPRELPLPSIRSAVESINELN--NEQNNYFEGCYLHOEGDCPIPKRRFF 1549
                |||               :||
Db       2883  OEFSAETOEKDAOLHLDQOVKEIINSIQANTDNEVDNAKTSGLNNTTEREPYNNK-- 2940
                |||               :||
OY      1550  KLMSGLSLAHSEISHKTRAKINKSVVDLIYDKENIDRSNRNFSDVLLDLHAERIKFA 1609
                |||               :||
Db       2941  ---KNAILKLYVDSQTQEALIN-----GYDATEDELQEAANSKLKILLDKKKQIGLA 2990
                |||               :||
OY      1610  --NGELIDECAPNKHEMK 1625
                |||               :||
Db       2991  HTNNEVDIDYNEVSOKMK 3008

RESULT 9
US-09-336-447A-76
Sequence 76, Application US/09336447A
Patent No. 6310190
GENERAL INFORMATION:
APPLICANT: HANSEN, ERIC J.
APPLICANT: AEBI, CHRISTOPH
APPLICANT: COPE, LESLIE D.
APPLICANT: MACIVER, ISOBEL
APPLICANT: FISKE, MICHAEL J.
APPLICANT: FRDENBURG, ROSS A.
TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
FILE REFERENCE: AMCY:024
CURRENT APPLICATION NUMBER: US/09/336,447A
CURRENT FILING DATE: 1999-06-21
NUMBER OF SEQ ID NOS: 98
SOFTWARE: Patentin Ver. 2.1
Seq ID NO 76
LENGTH: 3788
TYPE: PRT
ORGANISM: Moraxella catarrhalis
FEATURE:
NAME/KEY: MOD.RES
LOCATION: (1036)..(3786)
OTHER INFORMATION: Xaa = any
US-09-336-447A-76

Query Match          3.2%; Score 280; DB 4: Length 3788;
Best Local Similarity 18.2%; Pred. MO.2.5e-08;
Matches 308; Conservative 285; Mismatches 700; Indels 396; Gaps 79.

OY      29     ENSPTKPLNPRTNKSPLDPNSSDYTSSEDOCKGEKKKDTAFOTSFDNPFDLDSNI 88
                |||               :||
Db       1132  YGSKSEQIDNPTEINILINEARNACKITRKMAROGENGSEQIDNP--RFEINILINEARNAR 1189
                |||               :||
OY      89     -DIQOTTQHQQOOPOOQOOLSQTDNNNLIDERSFQTPMTSTILDITLKONP--TVDKVNF-- 142
                |||               :||
Db       1190  IGDIGRAKVSEQIDNPRTTEINILINEARNMALENXVEEGSEQID----NPRTTEINILINEAR 1245
                |||               :||
OY      143  NHAPTYINTSPKMSLMKATKPASKKYAFVTYTNPEIHNYPDNRVVEEDQSOQKEDSVP 202
                |||               :||
Db       1246  NAXANY-----ATPSITTNALESNV-----EBGLXLXLSSEQIDN 1279
                |||               :||
OY      203  PLIQHQWQDPSQFNYSDBEDTNA SVPPPLPLHTTKPTFAQLLNKNEVNSEPALTDMLK 262

```



```
Db 1280 P-----RTEINLINEARNAKANYATPST-----INSLENGESEQIDNPR-- 1321
Qy 263 RENFSLNLSDEKYNLXLSPTNNNSKNVSDMSHLONLQASKNKTNNENHNSLFALKAP 322
Db 1322 -----TEINLI-----NEARNASTLTDGKXVSEQIDNPREINLINEARNAKXAN 1365
Qy 323 KNDIENPLNSTNADISLRSSGSSQSLQSLRND-NRVLSEVGPSPKVPKVGSLNDGIR 381
Db 1366 YATPSTINSITDLGTIVDFXXXSEQIDNPREINLINEARNAKANYATPSTINSIT 1425
Qy 382 GRSDEVEESLPRDLSRDLKLETTKEHADAPHNNEFTDAKSTVTKGOLLVSSDDLDSF 441
Db 1426 DLG-TIVDS---EQIDNPREINLINEA-----RNAVDALXTKYNALDXKVNSDXTSEQI 1476
Qy 442 DRSYNTFEOSIINLNSASQS-----QISLALEKQRTQDOEQQAQAE-----PEEFT 490
Db 1477 DNP-RTE---INLINEARNAKANYATPSTINSILAEOQLNGKTLTPVSEQIDNPRETE- 1530
Qy 491 SFSDNIKVKOEKSNLEFVVTIKKEPVSAITEIKAPKREPS-----SRIL 535
Db 1531 -----INLINEARNAKHDAASTEKGMDSEQIDNPRETEINLINEARNAALLESVEEGL 1585
Qy 536 RIKNEDEIAPADIHPKE-----NEANSVEDTDLAKKALNDESDTQNSTKM--SI 589
Db 1586 DLGSEQID-----NPRETEINLINEARNAQNT--LIEKTANKSEQIDNPRETEINLINEA 1638
Qy 590 RPHIDSDMKLEDSNDQDREND---DISREKSDILINDVQSOTSDITGDKYNGSSSIT 645
Db 1639 RNAID---KNEYSIKSEQIDNPRETEINLINEARNASTLTDGTRKSEQI---DNPRETEIN 1691
Qy 646 KTLAPPRSDNDKENSLEDP-----ANNESL-OOOLEY-----PHTKEDD 686
Db 1692 INARNAQNTLLEKSEQIDNPRETEINLINEARNAALHEOQLELTYSQIDNPRETEIN 1751
Qy 687 SILANSSNAPPELTLPVYEANDYSSFNQVYTFDAYSSFEESLSREHETDS--KPI 744
Db 1752 INEARNASSDSQIDNPRETEINLINEARNAKKA-DADASF-ETLTSEQIDNPRETEIN 1809
Qy 745 ISLIMHQ-----EKOKKHQIHVPPK-QIIASQYKKNQOESR---VTSDKYVIPAIO 794
Db 1810 INARNAFAATAIAKDSQIDNPRETEINLINEARNAKASSENTQIAKSEQIDNPRETE- 1867
Qy 795 FKFEKVVNVSRRVSPDMODLVNQSPLPELSE--DSGFOLDNFANYSNNTNRPSTPL 852
Db 1868 -----EINL-----INEARNAILLDOKSEQIDNPRETEINLINEARNA--AA 1910
Qy 853 STKNVLS-NIDNDPNVVEPEPEKSYAEIRNARLSANKAPN-QAPPLPPOPOSSTRS 909
Db 1911 ITKNGXSEQIDN-----PRETEINLINEARNAKAXAANDRSQIDNPRETEINLIN 1961
Qy 910 NSKKRVSRFPVPTFEIRTSALAPCDMYNDIDDFGAGSKPTIKAGMKTLSMDKDV 969
Db 1962 EARNAXANYATPSTINSNOADIA--ONQTDI-ODLAAYNE--LOSQIDNPRETEINLI 2015
Qy 970 KRLNKKGVTOEYIAKLVQDKPKKNSIVTDPEDRYEELQQTASIHNAITDSSIGRP 1029
Db 2016 NEARNAQDIANNINNIYELAQOQDSQIDNPRETEINLINEARNAVNERQTEAT---- 2071
Qy 1030 DSISTDMLPYLDELKPPATALLSADRLFMEOEVHPLRSNSVLYHPGAGATNSSM--- 1085
Db 2072 DALN-----SQIDNPRETEI--NLINAR-----NAIL---GDTAIVSNGSQSEQ 2111
Qy 1086 --LPEPDFELINSPARVNSNDVA-----ISGNASTISENOLDMNFDOO--ATIGOK 1135
Db 2112 IDNPRETEINLINE-ARNAKALLESVEEGLDLDSGRSEQIDNPRETEINLINEARNAAL 2170
Qy 1136 IOBOPKSAANTYRGDDGLASAPETPRPTKKESISSKPAKLSASAPKSPIKIGSPVR 1195
Db 2171 VEEGLLELSTRTI---DQSGSEQIDNPR-----EINLINEARNAQAIANNIN 2216
Qy 1196 VIRKNGSIAGIEDLPKATHPKKSPQGNESINHKVRDGGISPSGSGHOOHNPAM----- 1250
```

```
Db 2217 XIYE-----LAQOQDOKS---EQIDNPRETEINLINEARNAKANYATPSTINNQ 2262
Qy 1251 VSPSQYTD-----ATSTVDEKNDKVOHKPREKQKHHNRH--HHHHHOKTIDPQVD 1303
Db 2263 ADIAQNOTDIODLAAYNELQSEQID---NPRETEINLINEARNAATHOYNEKQTASQID 2319
Qy 1304 DEIPDVLGREGKLFEEVVLGIKINLINLPDINTHKGRFTLTDNGHCVTTEPYNM--DDHN 1361
Db 2320 NPRETEIN-----INEARNAKASSENTQIAKSEQIDN-----PRETEINLINEARN 2365
Qy 1362 VAIGKEFELTVADSLFTLTK-----ASYEKPRCTLVEYTEKKVKSRRNLSRLGSKDI 1417
Db 2366 AMILGDPAL-VSNQDOKTOLKEFKYSEQIDNPR-----TEINLI--NEARNAAGCTII 2415
Qy 1418 ITTKFVPT-----EVKDTYANKFAPDGSFARCYIDLOQFEQITGKASQF 1463
Db 2416 PLDDXXPSSEQIDNPRETEINLINEARNAKANYATPS-----TTNSLHEOQLGKSEQI 2470
Qy 1464 D-----LNCFNEMETWMSNGQPMKRGKPYKIAOLEVMKLYVPRSDPREILPTSRSAYES 1518
Db 2471 DNPRETEINLINEARNAKAXAN-----YATPSTINIFPKXGSEQIDNPRETEINLINEA 2521
Qy 1519 IN-----EINQONNYFEGYLOEGGDCPIFKRFFKLGTSLLAHSEISHKTRA 1568
Db 2522 RNAKANYATPSTINNINNIYE-----LAQOQDQHSSDI 2556
Qy 1569 KINLSKYVD 1577
Db 2557 K-TLSEQID 2564
```

```
RESULT 10
US-09-134-001C-5106
; Sequence 5106, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064, 964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5106
; LENGTH: 1177
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5106
```

Query Match 3.2%; Score 274; DB 4; Length 1177;
Best Local Similarity 19.6%; Pred. No. 1.2e-08;
Matches 218; Conservative 169; Mismatches 351; Indels 374; Gaps 54;

```
Qy 69 KKTAFQTSDFDRNFDLNSIDIOQTIOHQOQPPQOQOOLSOTNNLI-----DEFS----- 119
Db 3 KKGSVIMSWFEDKLCGDNGSN-DLLLRKNKRROSOOSKQNNQDLSLPPONNDIYSRPGK 61
Qy 120 FQTPM-TSTLDLTQONTLVQDVNENHAPTYINTSPKNSIMKATPKKSPKVAATVTPNE 178
Db 62 FRPPIQVSENYETOKN--ENYNEH-----NOEFTNDIMSYOH-----DNPE 102
Qy 179 I-----HHYEDNVEEDOS--OQK-----DSVEPPLIOHMKDPSQFVNSDE 220
Db 103 FDSGSKHRRRRQVYKSHDQSKITQKQFADNNTYNNNSV--FNQNDKKSXGQKRSIQ 159
Qy 221 DTNASVPTPLPTTKTPFAOLN--KNEYVNSEPALTDKLRKRNFSNLSLDEKYNLY 278
Db 160 SENIKRANKPTKNTSTSEFTYLNHSFKSSEV---PSAIFGKKRRPTIENGVIPEPKEL- 215
```

QY 279 LSPINNNSKVSMD--SHLONLQDASKNKNTNENIHNLSFALKAPKNDIENPLSLTNA 336
 Db 216 -----NDKEVQDEVSHTKSIDASKVNSNDNNIE-----KNQ 251
 QY 337 DISLRSSSSOSLSLNDNRVLESVPGSKVNPGLSLNGIKG-----ESDEVESLL 392
 Db 252 QKQOQTQOTESSSSNNHNN---VERSNVQTKRKTPNISKVNTNINENIYASQIVE--- 305
 QY 393 PRDLSRDLKLETTKEHDAEHNENENFIDAKSTNTNGQLVSSDDHLU--SFDSY-----N 446
 Db 306 -----EIRREKERVVLQKRRF--KKALQOKRQONQOSEDSIQALDEMYAKQAOH 354
 QY 447 HTEQSILULNLASASQSLSLALEKQROTQOEQTOAEPBEETSPSDNIKVKQ-----EP 502
 Db 355 YTGESSLDENESNODS--SNSLEKQ-----NNSNIDNKEKQONNTP 395
 QY 503 KSNLEFVVTIKKEPVSAETIKAPREFSSRLIRIKNEDEIAPADIHKKKEANSIHE 562
 Db 396 LFNTF-----EIDLTTSQVKKVNEETSK-- 421
 QY 563 DTDALLKALNDEESDTTONSTKMSIRPHIDSDWKLSDN-----DGDREDNDISREFK 618
 Db 422 -----NDED-----LVSSNHYHSDDAEVEDAEYHELDNRQONQNS--Q 460
 QY 619 SDILNDVQSDIIGDKYGNSSSETTTKTAPRSDNNDKENSKELEDPANNEISLQOOLE 678
 Db 461 DDILSSKSTSNM-----YDNALIS-----ASVDNNTERAKSNED-----KNDIE 499
 QY 679 VPH-----TKEDDSILANSSNIAPPELTLPVEANDVSSFNDDTKTFDAYSFSESL 732
 Db 500 ITHLDGTTSAKVSDKISNTNHNLEQDKNVKLKVNLSKSNST-----GQTR 549
 QY 733 REHETDSKPIFISIMHKOQKQKHQIHK-----VPTKOIIASYQOYKNEQESRVTSD 785
 Db 550 KQREGSGSRPFN-VLMTPSDKKRMQDNHKKVSVPELKPEKQANNAHKKRQSESKS---- 603
 QY 786 KVKIPNAIOFKKFKRVNVMSRRVSPMD-----DLNVS-----QFLPEL 825
 Db 604 -----EEFKQINT-NRETDNSYESNGIEHDMNSSDEHAYETPSSKQODQIOQL 652
 QY 826 SEDSGFKDLNFANYSNNTNRPSEFTPLSTKNVLSNIDNDPNVEPEPEKSYAEIRNARL 885
 Db 653 ODDFHEFENAHAKI-NNSN-----ETGNGNSINHSKR- 683
 QY 886 SANKAAPQAPPLPQRPQSSSTRSNSKRVSR---FRVPTFEIRRTSSALAPCDMYNDIF 942
 Db 684 --SOYSTNESKNIDQTSNNSSTSNQNFQIRKGPNIKLPYOL--LEAPEHERQDQWI 738
 QY 943 DD-----FGAGSKPTIKAGKMTLPMDK-----DDVKRL 973
 Db 739 DNKKOELMDALYFNVPAEAKNVTEG---PSVTRFELSEKGVSVSRITLALODDIKMAL 794
 QY 974 NAK-----KGYTODEYINAKLVQDKPKR--NSIVTDPEDRYEELQOTAS---IHNATI 1021
 Db 795 AAKDIRIAPIPGTSVLQIEVPNQNPVKVNLRSIIESPFRKNTESKLYVAMQYRINNE-- 852
 QY 1022 DSSIYGRPDSISTDMLPYLSDELKAPPALLS 1053
 Db 853 -----PLLMD-IAKTPHALIA 867

RESULT 11
 US-08-603-753D-2
 : Sequence 2, Application US/08603753D
 : Patent No. 5891857
 : GENERAL INFORMATION:
 : APPLICANT: HOLT, JEFFREY T.
 : APPLICANT: JENSEN, ROY A.
 : APPLICANT: PAGE, DAVID L.
 : APPLICANT: KING, MARY-CLAIRE
 : APPLICANT: SZABO, CSILLA I.
 : APPLICANT: JETTON, THOMAS L.
 : APPLICANT: ROBINSON-BENION, CHERYL L.

APPLICANT: THOMPSON, MARILYN E.
 TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
 TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
 TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
 NUMBER OF SEQUENCES: 29
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: ARLES A. TAYLOR, JR.
 STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
 CITY: DURHAM
 STATE: NORTH CAROLINA
 COUNTRY: USA
 ZIP: 27707
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
 COMPUTER: IBM PC/XT/AT compatible
 OPERATING SYSTEM: Windows 3.1
 SOFTWARE: WORD PERFECT 6.1 and ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/603,753D
 FILING DATE: 20 FEB 1996
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: U.S. 08/373,799
 FILING DATE: 17 JAN 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: ARLES A. TAYLOR, JR.
 REGISTRATION NUMBER: 39,395
 REFERENCE/DOCKET NUMBER: 1242/2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919) 493-8000
 TELEFAX: (919) 419-0383
 TELEX:
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1863
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 HYPOTHETICAL: no
 ANTI-SENSE: no
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 INDIVIDUAL ISOLATE:
 DEVELOPMENTAL STAGE: adult
 TISSUE TYPE: female breast
 CELL TYPE: ductal carcinoma in situ, invasive
 CELL LINE: breast cancer and normal breast tissue
 CELL LINE: not derived from a cell line
 ORGANELLE: no
 IMMEDIATE SOURCE:
 LIBRARY: cDNA library derived from human
 CLONE: obtained using published sequence
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT: unknown
 MAP POSITION: unknown
 UNITS: unknown
 FEATURE:
 NAME/KEY: BRCA1 protein
 LOCATION: 1 to 1863
 IDENTIFICATION METHOD: observation of mRNA and
 IDENTIFICATION METHOD: antisense inhibition of BRCA1 gene
 OTHER INFORMATION: BRCA1 protein has a negative
 OTHER INFORMATION: regulatory effect on growth of human mammary cells.
 PUBLICATION INFORMATION:
 AUTHORS: MIKI, Y., et. al.
 TITLE: A strong candidate gene for the breast and
 TITLE: ovarian cancer susceptibility gene BRCA1.
 JOURNAL: Science
 VOLUME: 266
 PAGES: 66-71
 DATE: 1994

RELEVANT RESIDUES IN SEQ ID NO: 2: grainin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 1214-1223
us-08-603-753D-2

Query Match 3.1%; Score 270.5; DB 2; Length 1863;
Best Local Similarity 19.1%; Pred. No. 3.6e-08;
Matches 339; Conservative 239; Mismatches 610; Indels 591; Gaps 83;

18 QLOPSSASIFNSPTKPLNPRNTSKPLSDNSSD-----TYSSEOQEKKEKKKT 72
Db 170 RIOPKTSYI-----ELGSDSEDIYKAKATYGVGQELLQITPQGT 212
Qy 73 AFQTFDR-----NEDLNSIDIOOTIQH-----OQPPQOQOOLSQTN 112
Db 213 REIISLDSAKKACER-----SEFDVINTENHOPSNNDLNTERKRAERHEKQV--SSVN 268
Qy 113 NLIDFSPQTPMT-----STLDLTKQNPYDKV--NENHAPYINTSPNKSIMKAT- 162
Db 269 LAVEPCGTTHASLQHESSLLTFKDRMNVKEAEFCNKSQGLARSOHNWAGSKETC 328
Qy 163 -PKASPKKAFVTVNPEIHHPDNRYEEDQSOQK-----EDSVEPPLIQ----- 206
Db 329 NDRRTPS-----TEKKVDLADPLCEERKEMNOKLPCSENPDTEDVWITLNSIQV 382
Y 207 HOMKDPQSFNYSDE-----DINASVPPTPLHTTKPTFAQLLNKNVNSEPE 254
Db 383 NEM-----FSRDELLGSDSHDSESNAKV-----ADVLDVLEW----- 419
Qy 225 ALTDMLKRENFNSLDEKVNLYLSPTNNNNSKNVSDMSHLQNLQDASKNTNENIH 314
Db 420 -----DEYSGSS--EKIDLAS-----DPH-----EALICK-SDRVHS 449
Qy 315 LSFALKAPKNDIENPLNLTNADLSRSSGSSQSLQSLRNDNRVLESVPGRKKVNPGL 374
Db 450 -----KSVESDIEDKI-----RGKTYRKASLPLNLSHVENLI----- 482
Qy 375 SLNDGKGFSE-----VSESLPRDLSRDLKLETKEDHAPENENITDAKSTNKGGL 431
Db 483 -----IGAVSEPOITIOERPLTNKLRKRPTSGILHP-----EDT-----KKAOLA 524
Qy 432 VSSDHLDFDRSYNTEOS--ILNLINSASQISLMALEKQ-----QTOBOEOT--Q 482
Db 525 VQKTEPM--INQGTQTEONGQVMTNNGHENKTKGDSIQNEKNPNPLESLEKESAFKT 582
Qy 483 AAE-----EETSFSQNIKYQOEKPN-----LEPVKTIKEPVSATE 522
Db 583 KAEPISSISINLELIMHNSKA--PKNRLRKSSTRHIALELV--VSRNLSPNCTE 638
Qy 523 IKAPKEFSRILRIKNE-----EIAEPADIHPK--KENEASHVEDDAL 568
Db 639 LQIDSSSESEIKKKKYNNPVRHNSRLQMEKKEPATGAKKNKNEOTSKRHSDTTP 698
Qy 569 KALANDESDTTONSTKSIHF-----HISDMKLED--SNDQREDNDI-----SR 615
Db 699 ELKLTNAPGSFTKCSNTSELEKFNPSLPREEKEKLETVKSNNA--EDPKLMLSGER 756
Qy 616 FEKSDILNVQSDTIIGKYG--NSSSETTKTLAPPSDNDND-----KENSKL-- 664
Db 757 VLQTEKSVSSSISLVPGTDYQTESISLELVSTLGAKEPCKVCYQCAAFENPGLIH 816
Qy 665 ---EDPAN--ESLQOL--EVPHTKEDDSLANSNSIAPPEELLPLVPEANDYSFN-- 715
Db 817 GSKDRNDTEGFKYPLGHEVNSRFTSIEMSESLDAQLOLNTFVYSKQSGAPSPNG 876
Qy 716 ---DVTKTDAYSSFEESISREHETSKPLNFIStHKKOEKKQ--IHKVPTKQIIAS 770
Db 877 NAEBCATPSAHSG--SLKKQ--SPKVTF--ECBQKEBNQCKNESNIKIPQVQTVNITAG 928
Qy 771 YQOYKNEOSRYTSDKVKIIPNAIQKKFKFVNNMSRRVYSPDDDLNVSGF--LPELSEDS 829
Db 929 F--PVVGQKRPVDMNAKCSIKGSRFCLSSQFQENETGLTTPNKHGLQNLQYRILPPLPK 987
Qy 830 GF-----KDL--NFANYS-----NMTNRPGRSFTPLSTKAVLSINIDPNVVEPPEP 873

Db 988 SEVTKCKKNLLEENFEHSHMSPEREMGENENIPSTVSTISRNNIRENVKE----- 1038
Qy 874 KSYAELRNARRLSANKAARQAPPLPPQOPSSSTRNSKNKVRFRVPEIRTRTSALA 933
Db 1039 -----ASSNINEVG-----STNEVGSSINEG 1062
Qy 934 PCMDYNDIFDDFGAGSKPTIKAGMKTLPSMDKDVKRILMAKGVTODEYINAKLVQK 993
Db 1063 SSD--ENIQAEIQRNNGPKLNA-----MLRGLVQPEVYKQSLPGSN 1102
Qy 994 PKNSIVTPDEPKYELQOTASIHNATIDSSYGRPDSITDMLPYL--SDELKKPPTALL 1052
Db 1103 CKHPEI--KQOEYEVQGT-----VNTDPSPYLISNDLEQ----- 1136
Qy 1053 SADRLMEQDEVHPLRSNVLVHPGAGATNSSMLPREDELINSPPARVNSNSDNVAISG 1112
Db 1137 -----MGSNHSQVCSFPPDIL-----DDEIKE 1161
Qy 1113 NASTISFNOLDMNFDDQATIGOKIOEQPASKS-----ANTVRGDDGLA 1156
Db 1162 DT---SFAENDIK--ESSAVFSKVQKGLSRSPFTHTHLAGYRGAKLLESSENL 1217
Qy 1157 SAEPTPTPTKKESSISKRAKLSSASPRKSPKIGSPVRYIKKNGSLAGIEPIPKATHKP 1216
Db 1218 S--EDELPCQFHLLFGKVNIQSGTRHSTV-----ATECLSKNTEEN 1259
Qy 1217 KKSFGQ--NEISNAKYRQDGISPSGSEHOQHNPVNSVP--SQYTDATSTVPENKDVQH 1273
Db 1260 LSLKNSLNDSCSNQVILAKASQEHNLSEETKCSALFSGSCSELEDLTANTNTQDPELLG 1319
Qy 1274 KPREKQKQNNHNNHNNHNNQ--KTDPGVYDDELPRVGLQERKGLFFVLGIKNINLP 1330
Db 1320 SKQMR-----HOSQGVGLSKDELVSDBERGIGLENNQ----- 1356
Qy 1331 DINTHKGRTLLDNGHCVTTPRYNMDH--NVAIGKEPELTVADSLFILTAKASYEK 1388
Db 1357 -----EESQMSNLGEAASGCESETSVSD----- 1381
Qy 1389 PGCTLVEVTEKKVYKSRNRLSRFGSKDITTTKFPVTEVKQDWANKFADGSPARCYID 1448
Db 1382 -----CSGLSOSDILT-----TQORDTMQNHNL--T-----TK 1406
Qy 1449 LOOFEOITGKASOFDLCNENEMETMNGNQPM-----KRGRPYKIAOLEVMKLYPRS 1502
Db 1407 LQEBMALEAVLEQHGSPNSSTPILSDSALDELRLNPQDSISEKYLQISQSSSEFPIS 1466
Qy 1503 -DREILPTIRSAYESINELNBNQNNYFEGYLHOEGDCPIFKKRFKIMGTSLAHSE 1561
Db 1467 QNEGXSADKFEVSADSTSKNKE-----PGEVRSFSPKCPSLDDRMYHMSCGSLQNRN 1521
Qy 1562 ISHKTAKTNLSKVYDLYVDKENIDRNSNHRNSDVL 1600
Db 1522 YPQOE--ELIKYVD--VEEQULESGPHDLTETSYL 1553
RESULT 12
US-09-099-753-2
Sequence 2, Application US/09099753
Patent No. 6149903
GENERAL INFORMATION:
APPLICANT: HOLT, JEFFREY T.
APPLICANT: JENSEN, ROY A.
APPLICANT: PAGE, DAVID L.
APPLICANT: KING, MARY-CLAIRE
APPLICANT: SZABO, STILLA I.
APPLICANT: JETTON, THOMAS L.
APPLICANT: ROBINSON-BENION, CHERYL L.
APPLICANT: THOMPSON, MARILYN E.
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:
ADDRESSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 Inch, 800 KB storage
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,753
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/603,753
FILING DATE: 20 FEB 1996
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: ductal carcinoma in situ, invasive
CELL LINE: not derived from a cell line
ORGANELLE: no
IMMEDIATE SOURCE:
LIBRARY: cDNA library derived from human
CLONE: obtained using published sequence
POSITION IN GENOME:
CHROMOSOME/SEGMENT: unknown
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: BRCA1 protein
LOCATION: 1 to 1863
IDENTIFICATION METHOD: observation of mRNA and
OTHER INFORMATION: BRCA1 protein has a negative
PUBLICATION INFORMATION: regulatory effect on growth of human mammary cells.
AUTHORS: Miki, Y., et. al.
TITLE: A strong candidate gene for the breast and
TITLE: ovarian cancer susceptibility gene BRCA1.
JOURNAL: Science
VOLUME: 266
PAGES: 66-71
DATE: 1994
RELEVANT RESIDUES IN SEQ ID NO: 2: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 1214-1223
US-09-099-753-2

Query Match 3.1%, Score 270.5; DB 4; Length 1863;
Best Local Similarity 19.1%; Pred. No. 3,66-08;
Matches 339; Conservative 239; Mismatches 610; Indels 591; Gaps 83;
QY 18 OLPOSSASIFNSPTKPLNPRNKSPLDNPSSD-----YTSEDOQKGEKKOT 72
DB 170 RIQPKTSYVI-----ELGSDSSSEDTVKNATYCSVDOELQITTPGT 212
QY 73 AFOTSFDR-----NFDLNSIDIOQTIOH-----QOPOQOQOQISQDN 112
DB 213 RDEISLDKAKACEF-----SETDYNTHEHQPNSNDLITTEKRAERPEKYQG--SSVSN 268
QY 113 NLIDEFSFQTPYT-----STLDLTKQNPYDKV---NENHAPTYINTSPKSIKKAT- 162
DB 269 LHVEPCGJTNHASSLOHENSILLTKDMMNVEKAFCFKSKOPGLARQSHRNWAGSKETC 328
QY 163 -PKASPKVAFVTPNPELHHYPDNVEEDOSQK-----EDSVEPRLQ----- 206
DB 329 NDRRTPS-----TEKKVDLADPLCEKKEWKKLPKSENPREDVEDVPWITLANSIQV 382
QY 207 HOWKDPQSFNYSDE-----DTNASVPPPTLHTTKPFPALINKNNEVSBE 254
DB 383 NEM-----FSRDELGLGSDSDHDESESNKY-----ADVLDVLENY----- 419
QY 255 ALTDKLRKRENFSLDEKYNLYLPTNNNSKNVSDMSHLQMLDASKNKTENIHN 314
DB 420 -----DEYSGSS--EKIDLLAS-----DPH-----BALICK--SDRVHS 449
QY 315 LSFALKARKNDENPLNSLTNADISLRSSGSSQSLQSLRNDNRVLESVPKRVNGL 374
DB 450 -----KSVESDIEDKI-----FGKYRKASLPNLSHVENLI----- 482
QY 375 SLNDGKGFSDP---VVESLPRDLSPDKLETTKEHDAPEHNENFIDAKSTNTKGGL 431
DB 483 -----IGAFVSPQIIOGRPLTNKLRKRRTSGIAP-----EDFI-----KKAOLA 524
QY 432 VSSDHLDSFDRSYNHTQDS--ILNLNSASOSQISLNALEKQ-----QTOEOEQT--Q 482
DB 525 VQKTPEM--INGTQOTEDQNGVNMNTNSGHNKTKGDSIQNEKPNPIESLEKESAKET 582
QY 483 AAE-----EETSFSDNIRVKQEPKSN-----LEFVVTIKKEVSATE 522
DB 583 KAEPISSSISNLELNIHNSKA---PKNRLRRKSSRTHIALVLY--VSRLSLSPNCTE 638
QY 533 IKAPKREFSSRLIRIKNED-----EIAEPADIHPK--KENANSHEVEDTALL 568
DB 639 IQIDSCSSSEELIKKKKYNMOPVHRNRNQLMGEKREPARGAKKSNKPNOTSKRHSDFRP 698
QY 569 KALNDEBESDTTONSTKMSIRF-----HIDSMKLED--SNDGREDNDIT----SR 615
DB 699 ELKLTNAGSFTEKCSNTSELKEFVNPSLPREKEKEKLETVKVSNA--EDPDLMLSGER 756
QY 616 FEKSDIINDVSGTSDIIGDKYG--NSSEIITTKTLAPRSDND-----KENSLSL-- 664
DB 757 VLQTERSVESISSLVPGTDGTQESISLLEVSTLGAKKTEKNCVSCCAFAENKGLIH 816
QY 665 ---EDPANN--ESLQOOL---EVPHTKEDDSILANSINIAPEBELTLPVVEANDYSSEN--- 715
DB 817 GCSKRNNDTEGFKYPLGHEVNHRETSTIEMESLDAQYLQNTTKYVSKROSFAFSMPG 876
QY 716 ---DVTKTFDAYSSFEESI--REHETDSKPIINFISIMHKQOKKHQ--IKHVPYKQIIAS 770
DB 877 NAEBCATFSAHSG---SLKQK---SPKVPF--ECEQKEENGKNESNIKPVQTVNITAG 928
QY 771 YQKYNEDESKYTSOKVAILPNAILOFKKREVVNMRVYSPMDLNLNSQF--LPLSLSDS 829
DB 929 F--PVYGQKDPYDANKCSIKGSRFLCSQFGNFTGLTPKAGLLQNLNPRIPLPEPIK 987
QY 830 GF-----KDL---NFANYS-----NNTNRPSTPLSTKNVLSINDNDNPVPEPEP 873
DB 988 SFVTKCKKNLLEENFEBHSHSPERENGENIPSTVSTISRNINENYFKE----- 1038

```

0Y 874 KSYAEINABRRLANKAARQAPRLPPOROPSPSTSTNSNKKVRSPRVLPFELJRTSSALA 933
Db 1039 -----ASSNINEVS-----STNEGVSSINIG 1062
0Y 934 PCMDYNDIFDFGAGSKPTIKAEGMTLPESMOKDVKRILNKKGVTODEYIMAKLVDOK 993
Db 1063 SSD--ENIOAEIRGNKPKRNA-----MLRLGVLPREYVKOSLPQSN 1102
0Y 994 PKKMSIYTPDEDEYEELQOQASIAHNMTIDSSITGRDISITPMLYL-SDELRKPPALL 1052
Db 1103 CKHPEI--KQOEYEVVOT-----VNDPSPYLLISDNLEOP----- 1136
0Y 1053 SADRLEMEOEVRNHLRNSVLYVHNGAGATNSSLPEPDELINSAPRVNNSDVAISG 1112
Db 1137 -----MSSIASQVCSSTPDDL-----DDGEIKE 1161
0Y 1113 NASTISNOJDMNFDQATIGOKIOQOPAKS-----ANTVRGDDGGLA 1156
Db 1162 DT--SPRAENDIK-ESSAVFVKSVQGCILSRSPSPRTHTLAQGRGAKKLESSEEMLS 1217
0Y 1157 SAPRPTPKPKKSISIRAKLSASAPRKSPIKIGSPVAVIKKNSIAGIPIEPITAPHKP 1216
Db 1218 S--EDELPCFOHLLREKVVANITSOSTRSTV-----ATECLSKNTEEN 1259
Y 1217 KKSFOG--NEISNMKYVNDGDISPSSGSEHOQHNPSWVSP-SQYTDATSTVPDENKVOH 1273
Db 1260 LLSLKNLSDCSNOVILAKKSGEHNHLEETKCSASLIFSQCELEBDLTANTQOPFLIG 1319
0Y 1274 KPRKQOKKHNNHHNNHHNNQ--KNDIGVYDDELPRVGOERBKLEFRVGLGINKTLP 1330
Db 1320 SSKQMR-----HOSESQGVSLDKELVSDDERGTGLEENNO----- 1356
0Y 1331 DINHKGRFTLTDNGVHCVTTPREYMMDDH--NVAIGKEFELTVADSLFEILLTKASYEK 1388
Db 1357 -----BEQMDNSNLGSAAGCSEHSVEED----- 1381
0Y 1389 PRGTLVEYTEKKYVKSNNRLSRLFGSKDIIITTTKVEYVTKOTMANKARPDSFARCTID 1448
Db 1382 -----CSGLSSOSDILT--TOORPTOMHNL-----IK 1406
0Y 1449 LOQEPDITCKASQFDLNCENEMETWSNGNOMP-----KRCKPKIAOLEVKMILYPRS 1502
Db 1407 LOQEMAELEAVLEQHGSPNSIPSTIISDSSALEDLRNEQOSTSEKVLQTSKSSSEYIS 1466
0Y 1503 -DPEILLPTSRSAYESINELNNEQNNYEPGYLHQEGDCPIFKRKRFKLMQTSLLAHSE 1561
Db 1467 QNEPGXADKFEVSAUSDSTSKNKE-----PQVERSSPSKCPDLDRWMHNSGSGSLQRN 1521
0Y 1562 ISHKTAKINLSKVVDLYVDKENIDRSNHRNSDVL 1600
Db 1522 YPQOE---ELIKYVD--VEEOOLEBSGPHDLTERTSYL 1553

RESULT 13
US-08-986-106-2
Sequence 2, Application US/08986106
Patent No. 6177410
GENERAL INFORMATION:
APPLICANT: HOLT, JEFFREY T.
APPLICANT: JENSEN, ROY A.
APPLICANT: KING, MARY-CLAIRE
APPLICANT: STEINER, MITCHELL S.
APPLICANT: ROBINSON-BENION, CHERYL L.
APPLICANT: THOMPSON, MARILYN E.
TITLE OF INVENTION: THERAPEUTIC METHODS FOR
TITILE OF INVENTION: PROSTATE CANCER
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
STREET: BOULEVARD
CITY: DURHAM
STATE: NORTH CAROLINA

```

```

COUNTRY: USA
ZIP: 27707

COMPUTER READABLE FORM: diskette, 3.50 inch, 1.44MB storage
MEDIUM TYPE: IBM PC/Xt/AT compatible
COMPUTER: IBM PC/Xt/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/986,106
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/603,753
FILING DATE: 20 FEB 1996
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
PUBLICATION INFORMATION:
AUTHORS: MIKI, Y., et al.
TITLE: A strong candidate gene for the breast and
        ovarian cancer susceptibility gene
TITLE: BRCA1.
JOURNAL: Science
VOLUME: 266
PAGES: 66-71
DATE: 1994
RELEVANT RESIDUES IN SEQ ID NO: 2: grainin box domain
RELEVANT RESIDUES IN SEQ ID NO: at amino acids 1214-1223
US-08-986-106-2

Query Match          3.1%; Score 270.5; DB 4; Length 1863;
Best Local Similarity 19.1%; Pred. No. 3 6e-08;
Matches 339; Conservative 239; Mismatches 610; Indels 591; Gaps

QY   18 QLOPSSASATFSPTKPLNFPRTNSKPSLDPNSSD-----TYTSDQDEKKEEKDT 72
Db   :|||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
       RLOQPSTKVYI-----ELGDSSEEDTVANKATCSGDDELQITPGT 212

QY   73 AFQSPDR-----NPDLNSIDIQTQH-----QQQPPQQQLSOTDN 112
Db   :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
       RDELSDLSAKKAACEF---SETDVNTIEHQPSNNNDINTTKRAEHRHPKYQG--SSVSN 268

QY   113 NLIDEFSQPMT-----STLDLTKNQPTVDKY---NENHAPYINTSPNKSIMKKAT 162
Db   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
       LHVEPCGNTHASSLGHNSSLTLTKDRMANVEKAFCKSKSQGPLANSQNRNMGASETC 328

QY   269 LHVPCGNTHASSLGHNSSLTLTKDRMANVEKAFCKSKSQGPLANSQNRNMGASETC 328
Db   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   163 -PKASPKVATVYNPELIHHYPDKRVEEDOSQK-----EDSVPRPIQ----- 206
Db   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
       NDRKTPS-----TEKKVDLNADPLCEKEMKKQKLCPSENPRDTEVPWTTLNSSIQKV 382

QY   329 HQRKDPSQFNYSDE-----DTNASVPTPLHTTKPTFAOLLKKNNEVNSEPE 254
Db   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
       NEM-----FMSDELIGSDSHDSEBESNAKV-----ADVLYLVNV----- 419

QY   255 ALTDMLKRFENSLUDEKVNLYLPTNNNSKNVSDMSHLONLQDASKNKTNENIH 314
Db   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
       -----DEVSGSS--EKIDLLAS-----DPH-----EALICK--SDRVHS 449

QY   420 -----LSPALKAKRKNDLPNLNLTMAOISLSSGSSQSOSLGRNDNRVLVESVGSPKKVY 374
Db   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
       -----KSVESDIEDKI-----FGKTYKKKASLPULSHVTENLI----- 482

QY   375 SLNGIKKFQSPD-----VVESLLPRDLSRRKLETTETHEHAPBEHNENNFIADASTNTNKCOL 431

```

```

Db 483 -----IGAFVSEPOIIQIERPLTKLKRKRPTGLHP-----EDFL-----KRAOLA 524
Qy 432 VSSDDHLDSFDRSYNHTEOS--ILNLNSASQOISLMALEKOR-----OTOEOEOT--Q 482
Db 525 VQKTPRM--INQOTNOEQGYMNITNSGHEKTKGDSIQNKNRPPIESLEKESAFKT 582
Qy 483 AAEF-----EEETFSFNKIKKOEKPSN-----LEPVKTIKPEVATE 522
Db 583 KKEPISSISNELLEIMHNSKA---PKNRLLRKRKSTRHIALELY--VSRNLSPPNCTE 638
Qy 523 IKAPKREFSSRIILRIKNEO-----ELAEPADIHPK--KENANSHEEDDALL 568
Db 639 LQIDSCSSSEELIKKKYNOHPVHSHRNLOLMEKSEEPATGAKKSNKPREQSKRHSDTPE 638
Qy 569 KRALNDEESDTQNSTKMSIRF-----HIDSWKLED--SNQDREDNDNDI---SR 615
Db 699 ELKLTMAPSGFTCSMTSELKEFVNPSPREKEKELEYKVSNN--EDPKMLMSGER 756
Qy 616 FEKSDILNDVSDTITGDKYG--NSSSETTKTAPPRSDND-----KENSLSL-- 664
Db 757 VLQTERSVSSSISLVPCTDYGTOESISLLEYSTLGAKTEPKVCVSCAFAENPKGLIH 816
Qy 665 ----EDBANN--ESLQOOL--EVPRTKEDDSTLANSNIAPPEELTLVPEANDVSSFN--- 715
Db 817 GCSKORNDTEGKRYPLGHEVNHRSRTSIMESELDQOYLQTFEYVSKQSPAPSPNG 876
Qy 716 ---DYTKTDFVASSFEESISREHETDSKPIFISIMHKOEKQKHQ--IHKVPKQIIAS 770
Db 877 NAEBCATFSAHSG---SLKKQ---SPKYTF--ECEQKEGNCKNSNIKPVQTVITAG 928
Qy 771 YQOYKKEQSRVTSQVKYLPNAIQKFKFEVNVMSRRVSPMDLNVSOE--LPELSEDS 829
Db 929 F--PVYQKQKPVNMACSISKIGSRFCLSSQFRGNETGLITPNKHGILLQNPYRIPPLPK 987
Qy 830 GF-----FDL--NFANS-----NMTNRPSTPLSTKVLNSINDPVPVPEPPE 873
Db 988 SFYKTKCKNLLENFEEHSMSPREKMGENDIPSTVSTISRNIRENVKE----- 1038
Qy 874 KSYAETIRNARLSANKAANQAPPLPPOPOPSSTRSNKRRVSRFPVFEIRRTSSALA 933
Db 1039 -----ASSNINEVGS---STNEVGSISNEIG 1062
Qy 934 PCDMYNDIDDDGAGSKPTIKAGMTTLPSMDKDYKRIILNAKKYQODEYINAKLVDOK 993
Db 1063 SSD--ENIQAEIGRNGPKLMA-----MLRGLVLOPEVYKQSLPGSN 1102
Qy 994 PKNSIVTDPEDRYEELQOTASTIHNTIDSSIVGRPOSTIMLPLYL--SDELKKPPFALL 1052
Db 1103 CKHPEL---KKOEYEEVOT-----VNTDSPYLISDNLQEP----- 1136
Qy 1053 SADRLEMEQEVHPLRSNSVLYHPGAGATNSSMLPPEDELINSAPARNVSNNSDNYAISO 1112
Db 1137 -----MGSSHASQVCSFETPDDL-----DQBEIKE 1161
Qy 1113 NASTISFNOLDANFDQATIGQIKQDQPAKS-----ANTVARGDDGLA 1156
Db 1162 DT--SFAENDIK--ESSAIVFSKSVOKGELSRSPPTHTHLAGYRRGAKKLESSEENS 1217
Qy 1157 SAPETPRTPTKESISSAKLSSAPRKSPIKIGSPAVIKKNGSIAGIEPIPKATHPK 1216
Db 1218 S--EDELPCFQHLLEGKANNIPSOSTRHSTV-----ATECLSKNTEEN 1259
Qy 1217 KKSFOG--NEISNHKVRDGISPSGSEHQOHNPMVSV--SQYTDATSTVPDENNDVOH 1273
Db 1260 LSLKNSLSDCSNOVILAKASQEHLSSEFKCSASLFSQSCSELELDTANTNODPFLIG 1319
Qy 1274 KREKOKKHHHHHHHHHHK--KTIDIGVVDDELIDVGLQERGLKFFRVGLIKINMLP 1330
Db 1320 SSKOMR-----HQSQGVGLSDKDELVSDBDEERGGLLEENN----- 1356
Qy 1331 DINTHKGRTLTLDNGVHCVTTPEVNMDDH--NVALGKFEFLTVASLEFILTLLKASYEK 1388

```

```

Db 1357 -----EQSMDSNLGEAASCESETSVSD----- 1381
Qy 1389 PRGLIVEETKVKVKSNNRLSRFGSKDITTTKFPVTEKQDWMANKFAPDGSFARCYID 1448
Db 1382 -----CSGLSSOSDILT-----TOORDMOINL-----IK 1406
Qy 1449 LQGFQDQITGKASQFDLNCFNENETWSNGQPN-----KRGKPYIAOLEVMKLVVPS 1502
Db 1407 LQOEMAELEAVLQHGOSQPSNSYPSIISDSSALEDLRNPQOSTSEKVLQTSQKSEYPI 1466
Qy 1503 -DRELIPTSIRAYESINLNEQNNYFEGYLOHGQDQPIKRRFKPLMGLSLAHSE 1561
Db 1467 QNPEGKADKFEVSAUSSTSKNE-----PGERSSPSKCPSLDDRWYMHSCGSLQNN 1521
Qy 1562 ISHKTAKINLSKYVDLYDKENIDRSNHRNPSVILL 1600
Db 1522 YPQGE---ELIKYVD---VEBOQLSESGHDLTETSYL 1553

RESULT 14
US-09-007-678B-49
: Sequence 49, Application US/09007678B
: Patent No. 6342483
: GENERAL INFORMATION:
: APPLICANT: HOLT, JEFFREY T.
: APPLICANT: JENSEN, ROY A.
: APPLICANT: PAGE, DAVID L.
: APPLICANT: OBERMILLER, PATRICE S.
: APPLICANT: ROBINSON-BENION, CHERYL L.
: APPLICANT: THOMPSON, MARILYN E.
: TITLE OF INVENTION: METHOD FOR DETECTION AND TREATMENT OF BREAST CANCER
: FILE REFERENCE: Attorney Docket No. 6342483 1442-1-2-2
: CURRENT APPLICATION NUMBER: US/09/007,678B
: PRIOR FILING DATE: 1998-01-15
: PRIOR APPLICATION NUMBER: 08/373,799
: PRIOR FILING DATE: 1995-01-17
: PRIOR APPLICATION NUMBER: 08/182,961
: PRIOR FILING DATE: 1994-01-14
: NUMBER OF SEQ ID NOS: 61
: SOFTWARE: Microsoft Wordpad
: SEQ ID NO 49
: LENGTH: 1863
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1472)
: OTHER INFORMATION: Xaa-any amino acid
US-09-007-678B-49

Query Match 3.1%; Score 270.5; DB 4; Length 1863;
Best Local Similarity 19.1%; Pred. No. 3,6e-08;
Matches 339; Conservative 239; Mismatches 610; Indels 591; Gaps 83;

Qy 18 QLOPOSSASIFNSPTKPLNFPRTNSKPSLDPNSSD-----TYSQEOEGKEKKOT 72
Db 170 RIQPOKTSYVI-----ELGSDSSSDTYNKAITYCVGQDELQITPOGT 212
Qy 73 APQTSFDR-----NFDLNSIDIQOTIOH-----OOOQOQOQLOQOTN 112
Db 213 RDEISLDSAKKACER-----SETDVYTFEHHQPSNNDLNTTEKRAERHPEKTQG--SSVSN 268
Qy 113 NLIDFSPQTPMT-----STDLITKONFTYDKV--NENHAPTYINTSPNKSIMKAT-- 162
Db 269 LHYEPCGTMTNTHASSILOHENSILLTTRDMNVKEAKFCNKSQKGLARSGHNMWAGSKETC 328
Qy 163 -PKASPKKAAFTVYNPELTHYDPNRYEEDQSOQK-----EDSVYEPPLQ----- 206
Db 329 NDRKTPS-----TEKKVLDLMDPLCERKEWKKQKLPCCSENPRDIEDVPIITLNSIQKY 382
Qy 207 HOKKQPSQFNYSDE-----DTNAYVPPRPPLTLTKPTFAQLLNKNNEVNSEPE 254
Db 383 NEW-----FSRDELLGSDSDSHDGESESNAKY-----ADVLQVLENY----- 419

```

Qy	255	ALIDMUKARRNENSLUDEVNVNLYLSEPTNNNNNSKVNQMDSDHSLQLODASKNKNTENHN	314
Db	420	-----DEYGGSS--EKIDLLAS-----DPI-----BALICK--SDRVHS	449
Qy	315	LSFALAKPKNNDIENPLMLSLTNADISLRRSGSSOSLSLRNDRNRLVESVPSGPKVNVNGL	374
Db	450	-----KVEDIEDIKT-----FKYTKKASLPLNSHTYENLI-----	482
Qy	375	SLNDGKIGFSDE--VVESLLPRDLSRDKLETTKEHADAEHNNEFIKASTNTKCOLL	431
Db	483	-----IGAFVSEPOIIIOERPLTKLKRKRPRPSGLHP-----EDFI-----KKADLA	524
Qy	432	VSSDHLHDSRDRSNTHTEOS--ILNLLNSASOSLSLNALEOR-----QTOEOBOT--Q	482
Db	525	VOKTPEM--INOGTNOTEOGQVNMNTTNGHENKTKGDSIQNEKPNPDIESLEKSAKT	582
Qy	483	AAPD-----EETSPDNITKVKOEPKS-----LEFVYTIKKPEVSATE	522
Db	583	KAEFISSISSELELNIMHNSKA---PKKNRLRKSSSTHIALELY--VSRNLSPPNCTE	638
Qy	523	IKAKRREPSRILRIKNE-----ELIAPADIHPK--KENPANSHEVEDTALL	568
Db	639	LQIDSCSSSEIKKKKRYNQMPVRHNSNLODMGKEPATKAKSKNPNQDTSKRHDSTOFF	698
Qy	569	KKALNDEESDTONSTKMSIR-----HDSOMKLED--SDGDREDDNDT---SR	615
Db	699	ELKLTNAPGSFTKSNSTSELEKFEVNSJLPREEKEKLETVKYSNNA--EDPKDMLSGER	756
Qy	616	FEKSDILNDVSOQSDIIGDYK--NSSSELTITKTLAPPSQND-----KENSXL--	664
Db	757	VLOTERKVESSSISLVPGTDYGTQEBESILEVSTLQAKATEPKKCVSQCACAFENKGLIH	816
Qy	665	---EDPANN--ESLQOOL--EVPHTKEDDSILNSSNIAPPELTLPVEANDYSSFN--	715
Db	817	GCSKDNNDDEGFYPLGHEVHNHSRFTSIEMESELDQYLQNTFKYSRQSGFAFNSMG	876
Qy	716	---DYTFPDAVSSFEELSREHETSKPNTSINHKOQKQKHQ--IHKVPTKQIITAS	770
Db	877	NAEECATCFAHSG--SLAKO--SPKYTF--ECBOKEENOGKNESNIKPVQYVNTIYAG	928
Qy	771	YQCKNKOESRYTSDVKYKILNALIOFKFKFVNVMASRRVYNSPMDLDVNSQF--LEPLSDS	829
Db	929	F-PVYGOKQKPVDMAKSITIGSRFLCSQFQKNETGLITPKKHGLOLNPRIPLPEPIK	987
Qy	830	GF-----KDL--NFANS-----NNTNRPSPFTPLSTKNVLSNIDNDPNVVEPEP	873
Db	988	SFVTKCKKNLLEENFEHSHSPREMGNEGNIPSTVTSIRNNIRENFKE-----	1038
Qy	874	KSYAEIRNARRLSKANKAPQAPPLPPQROPSSTNSNSKRYRSRRRVFPETIRTSALA	933
Db	1039	-----ASSNINIEVGS--STNEVGSSINEIG	1062
Qy	934	PCDMYNDIIPDOFGAGSKPTKAGEMKTLPSMOKDVKRLNKKKQVTOEYINAKLVQOK	993
Db	1063	SSD--ENIQDELGNNGPKLNA-----MLRLGVLDPYVYKOSLPGSN	1102
Qy	994	PKKNSITVDPEDRLEEQLOQASIHATIDSSITGRPDSISTMLPYL--SDELKRPATLL	1052
Db	1103	CKHBEI---KKOEEVEVQI-----VMTDPSPILLSNDIEQ-----	1136
Qy	1053	SADRLFMEQEVHPLRSNSVLVHFGAGATNSMLPPEDFELINSPARVNSNSNDVAISG	1112
Db	1137	-----MOSHAQYCSBTDDLL-----DDGEIKE	1161
Qy	1113	NASTISNQLDMAFNDOATIGQKIQEOBPAKS-----ANTYRGDDGCLA	1156
Db	1162	DT---SFAENDIK--ESSAVNSKSVQKGELSRSPFTHTHLAQGYRGAKKLESSEEMLS	1217
Qy	1157	SAPETPTPKPKKSISSKPAKLSASAPRKSPIKISPRVAVIKKNSIAGIEPIKATHPK	1216
Db	1218	S--EDELPCQOHLLEKQVNNITPSQSTIRSHV-----ATBELSKNTEEN	1259

```

QY 1217 KKSQOG--NEJSHNKVADGGISPSSESEHOJHNSVAVP--SOYUATSTVDEKADQOH 1273
Dd 1260 LLSLAKNSLDCSNQVUILLAKSOZHNLSEETKCSASLSSQCELEDLTANTQDPLIG 1319
QY 1274 KPEKOKOKHHNNHHNNHHNKKO--KTJDIPGVYDDELIDVGLQERKLEFFRYLGIKINLP 1330
Dd 1320 SSKQMR-----HQSESQGVGLSDKELYSDERGJLEBNNQ----- 1356
QY 1331 DINTHKGRFTLLDNGVHCYTPPEYMDH--NVALGKFELTVADSLFILLTKASYEK 1388
Dd 1357 -----EEOGMDSNLGEAAGCSESTVSIED----- 1381
QY 1389 PRGTIVEVEKVKVSKSNRLSRFGSKDIIITTKFPRVCTYKANKFADPSFARCYID 1448
Dd 1382 -----CSGLSSOSDILT-----TOQRTMOHNL-----IK 1406
QY 1449 LQGFEDDITGKASQFDFLNCENEMETMSNCPM-----KROKPKYIAOLEKMYLPYRS 1502
Dd 1407 LQCEMALEAVLEQHGQSPNSYRPIISDSMALEDLRNPEGSTSEKVLQTSQKSEYIS 1466
QY 1503 -DPEELPLTISRASYESTINFLNNEQNNNYEBGUILHDEGGDCPLFKKRFKRLMGTSLAHSE 1561
Dd 1467 QNPRGXSADKFEVSADSSYTSKNE-----PQVERSSPSPKCPYLDIDRWYHNSCSEGLQARN 1521
QY 1562 ISHTRAKTNLSKVYDLIYVDEKIDNRSHNRNSDVLIL 1600
Dd 1522 YPPOE--ELIKVYD--VEEOOLESGPHDILTETSYL 1553

```

RESULT 15
 US-07-741-940-2
 Sequence 2, Application US/07741940
 Patent No. 5352775
 GENERAL INFORMATION:
 APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YOSUKE
 APPLICANT: THLIVERS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner, Birch, Mckie & Beckett
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/741,940
 FILING DATE: 19920109
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.035574
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2843 amino acids

TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-741-940-2

Query Match 3.1%; Score 270; DB 1; Length 2843;

Best Local Similarity 19.1%; Pred. No. 7e-08;
 Matches 287; Conservative 216; Mismatches 570; Indels 428; Gaps 64;

QY 22 OSSSASIFNSP---TKPLAFPRNTSKP-----SLDPSSSDYTSRODEKCK 66
 DB 924 RSSAAHTHNTNTYTKSESNRTCSMPYAKLEYKRSSNDLSNVSSNDGY -GKRGQMKRS 982
 QY 67 EEKKDTAQTSTF-----DRNFDDNSIDIOQTIOHOQ-----QQ 100
 DB 983 IESYSEDEDSKFCYQGYPADLAHKIHSANHMDNDGELDTPIYSLKYSDEQLNGRS 1042
 QY 101 PQOOOOLSTQDNNLIDFESFOTPMSTLDLTQONPTVQKNENHA-PTIINTSPKSIK 159
 DB 1043 PSQNEWAPKHIIEDEI-----KQEQROSRNOSTTYPVYTESTDDKHL-- 1087
 QY 160 KATPKASPKVAFTVYNPEIHYPDNRY-----FEEDSOOKEDSVF---PPLIOHOM 209
 DB 1088 KQPHFGQOECVPYRSKRANGSETRKVSNGHNGINOVQSQCQEDDYEDDKPTNYSKY 1147
 QY 210 KDPQFNVSDSDTNASVPPTPLHTTKPTFAQLLNKNNEVSEPEALTDMLKRENFSL 269
 DB 1148 SEEDHEEERPTNYSI-----KYNEKRHVDPIDYSLKYA--TDI 1187
 QY 270 SLDEKYNLTLSPNNNSKNVSDMOSHLODASKNTNEN-IHLSALKA--PKNDI 326
 DB 1188 PSSQKOSFSKSSGQSKTEHMSSTSTPSSNAKROHLPSSQOSNGOQKAA 1247
 QY 327 ENPLNSLTNADISLR-----SSGSSQSSLOSLRND--NRVLESYPGSPKKVNPGL 374
 DB 1248 TCKVSSINQETIQTCVEDPTPICFSCSSLSLSAEDICGNTQGEADSNANTLIQAI 1307
 QY 375 SLNDGKGFSDVEF---SLPRDISRKLETTEKHAPEHNNENF-IDAKSTNTNGQ 429
 DB 1308 KCKIGTSAEDVSEVPASQHPRTKSSRLQSSLSSESARHKAVFEPGAKSPSKGAQ 1367
 QY 430 LTVSSDDH-----LDSFRGYNHTQESLILNLNLSASQ-----ISLN 467
 DB 1368 TPKSPPEHVOETPLMFESRCTSVSSLSDFE-----SRISASSVOEPCSGVSGITSPS 1421
 QY 468 AL-EKQRTQEOEQTAEPSEETSESDNIKYOEKPSMLEFYKVIKKEPVSAFEIKAP 526
 DB 1422 DLPDSCGQTPMPSRSKTPPPQTA-----QTKREVPKN-----KAPTAEKRESGP 1467
 QY 527 KR-EFSSRLIRK-----NEDETAEPADIIHPKE 554
 DB 1468 KOAAVNAAVQRYVLPDADTLHFATESTPDGFCSSSLALSLEDPFOKDELRIIMP 1527
 QY 555 NEANGSHVEDTDLKKALMDEESDTTONSTKMSTRFHIDSDMKLEDSNDGREDNDIS 614
 DB 1528 VOENDNGNETSEQPKESNEOEKEAET-----IDSEKDLDDS-----DDDDIE 1573
 QY 615 REEKSDILNDVQSOTSDIIGDKYGNSSSEITTKTLAPPRSDNNDKENSKLEDPANNEISQ 674
 DB 1574 ILEEC-IISAMPTKSRKKKKAQATASKL-----PPVAKRPSQLPYVKLLPSQNR-IQ 1625
 QY 675 QOLEVPHRKEDP-----SILANSSNIAPPEEL----- 701
 DB 1626 POKHVSFTPGDMPRYVCVEGTPINFTATSLDLTIESPPNELAAGVGARGAOSGEFE 1685
 QY 702 ---TLPVYANYSFNDVTKTEDAASSFEESLSREHETDSKPIFISIMHKOEKOKKHQ 758
 DB 1686 KRDITLPTBERSTDEAGGKTSSV-TIPELDDNKAEBGDLLAECTIN-----SAMPKGSNR 1739
 QY 759 IHKVPTKQIIASVQ-----YKNE---QESRVTSDKVKIIPNAIOFKFKFEVNVMSR- 806
 DB 1740 PPRV--KKTMDVOQASASSAPNNKQDLGKKKKPTSPVKPIQNTFVTRVRKNADSKN 1797

QY 807 -----RVVSPMDDLNVNQFLPELSEDSGFKDLNFANYSNNTNR-----PSSEFPL 852
 DB 1798 NLNAERVFSDNKD--SKQNLKNNSKD--FND-----KLPPNEDRYGSAFSDPHHYPI 1849
 QY 853 -----STKNVLSINDNPNVVEPEPKSYAEIRAKRLSANKAAPNPAPLPQORQS 905
 DB 1850 EGTPEYCSFNDSLSLDFDDVDVLSREK--AEIRAKENKSEAKVTSHTBL-----T 1901
 QY 906 STRSNKRVSRFRVPTFEIRRTSSALAPCDMYNDIFDFGAGSPPTIKAEQMKTLPSMD 965
 DB 1902 SNOQSANKTQALAKQ---INR-----GQPKPILOQ--STFPQSS 1937
 QY 966 KODVKRILNAKKGYTODEYINAKLVQKP-----KNSIV--TD 1002
 DB 1938 KDIPDR-----GAATDEKIQNFALIENTPVCFSHNSLSLSDIDQENNNKENEPKETE 1991
 QY 1003 PEDRYE-----LOQTASHNATITSSITGRDSTIMPL 1040
 DB 1992 PPDQGEPSKPOAGYAPKSFVHEDTPVCFSHNSLSLSLIDSE-----DILLQECISSA 2046
 QY 1041 SDELKPPPTALLSADRLFMEOEVHPLRSNSVLVHPGAGATNSSMLPEPDFLINSAPRN 1100
 DB 2047 MKKKKKP-----SKLKQDNKHSFRNNGIL--GEDTLDLKDIDQRPDSEHGLSPD-- 2095
 QY 1101 VSNNSDNVAISGNASTISFNQDMNFDQATIGQKIQEQPASKSANTYVRGDDGLASAP 1160
 DB 2096 -SENFDMKAIQEGANSI-----VSLHQAAAACLSRQ--ASSDSDSLKSGISLQSP 2147
 QY 1161 TPTPTKRESISSKPAKUSSASPRKSPIKISPVRIKNGSIAGIEPIPKATHPKKSF 1220
 DB 2148 FILTPDQEE---KPE-TSNKGR-----ILKPGKSTLE--TKIESESKTI 2188
 QY 1221 QGNEISNHRVDRDGISPSGSEHQHNP-----SNVSPQOYTDTATSTVPDEN 1268
 DB 2189 KQKKVYKSLITGVKRSNEISGQMKQPLQANMPSISRGRTMIHLPGVANSSTSPYSK 2248
 QY 1269 K 1269
 DB 2249 K 2249

Search completed: March 17, 2003, 12:27:34
 Job time : 107.906 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:27:19 ; Search time 21.588 Seconds
(without alignments)
3552.777 Million cell updates/sec

Title: US-09-964-858-1
Perfect score: 8631
Sequence: 1 MNSYPSKLLPKDHSHLQLQ.....WVNLALQOQOQOQOQSSQO 1664

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	327	3.8	6281	10	US-09-815-242-12996
2	309.5	3.6	2437	10	US-09-815-242-5834
3	298	3.5	2025	10	US-09-815-242-5703
4	298	3.5	3158	10	US-09-815-242-12611
5	289.5	3.4	2344	10	US-09-815-242-12713
6	280	3.2	1139	9	US-09-820-843A-15
7	280	3.2	2368	10	US-09-815-242-5635
8	280	3.2	2368	10	US-09-815-242-12389
9	280	3.2	3788	9	US-09-052-267-76
10	276.5	3.2	1400	10	US-09-764-176-7
11	271	3.1	1616	9	US-09-820-843A-16
12	269	3.1	2843	9	US-09-987-482-1
13	268.5	3.1	1863	9	US-10-022-819-2
14	267.5	3.1	1863	9	US-09-734-672-4
15	267.5	3.1	1863	9	US-09-982-828-6
16	267	3.1	2434	10	US-09-815-242-5835
17	267	3.1	2478	10	US-09-815-242-5816
18	267	3.1	2478	10	US-09-815-242-12967
19	266.5	3.1	1111	10	US-09-815-242-12955

20	266.5	3.1	2843	8	US-08-681-219-32	Sequence 32, Appl
21	265.5	3.1	1863	9	US-09-734-672-2	Sequence 2, Appl1
22	265.5	3.1	1863	9	US-09-734-672-6	Sequence 6, Appl1
23	265.5	3.1	1863	9	US-09-982-828-2	Sequence 2, Appl1
24	265.5	3.1	1863	9	US-09-982-828-4	Sequence 4, Appl1
25	265.5	3.1	2285	10	US-09-932-183A-2	Sequence 2, Appl1
26	264.5	3.1	2076	10	US-09-815-242-5815	Sequence 5815, Ap
27	264.5	3.1	2186	10	US-09-815-242-12913	Sequence 12913, A
28	262	3.0	665	9	US-09-820-843A-107	Sequence 107, App
29	262	3.0	5795	10	US-09-815-242-12610	Sequence 12610, A
30	258	3.0	1884	10	US-09-785-770A-17	Sequence 17, Appl
31	258	3.0	1907	10	US-09-785-770A-16	Sequence 16, Appl
32	257.5	3.0	1478	10	US-09-801-368-52	Sequence 52, Appl
33	253.5	2.9	2086	10	US-09-815-242-5639	Sequence 5639, Ap
34	253	2.9	2665	10	US-09-864-761-34248	Sequence 34248, A
35	253	2.9	3256	10	US-09-919-172-98	Sequence 98, Appl
36	246.5	2.9	1421	10	US-09-924-154-13	Sequence 13, Appl
37	245.5	2.8	1285	10	US-09-982-091A-2	Sequence 2, Appl1
38	245.5	2.8	1805	9	US-09-820-843A-73	Sequence 73, Appl
39	241.5	2.8	1325	10	US-09-864-761-35612	Sequence 35612, A
40	240.5	2.8	1215	10	US-09-815-242-5908	Sequence 5908, Ap
41	240.5	2.8	1269	10	US-09-815-242-13113	Sequence 13113, A
42	239	2.8	2167	10	US-09-801-368-56	Sequence 56, Appl
43	238.5	2.8	1501	10	US-09-924-154-17	Sequence 17, Appl
44	236.5	2.7	1093	10	US-09-801-368-392	Sequence 392, App
45	236.5	2.7	1518	10	US-09-801-368-152	Sequence 152, App

ALIGNMENTS

RESULT 1
US-09-815-242-12996
Sequence 12996, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 12996
LENGTH: 6281
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12996
Query Match 3.8%; Score 327; DB 10; Length 6281;
Best local similarity 17.7%; Pred. No. 1.2e-08;

Matches 296; Conservative 279; Mismatches 607; Indels 488; Gaps 71;

```

QY 42 NSRPSLDPNSSDITYSEQ--DOEG-----KEEKD-----71
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2459 NGSFVITNADYTAATQAVRNMAETGLNGDNLATAKQAQKDALRQWTHLSDAQKOSITG 2518
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 -----TAFOTSFDRNPLDINSID-IDOTIOHQOQPOOQOOLSTDMNLIDESFOR 122
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2519 QIDSATQVTVQSVKDNATLNDNMANOLNRNSINKDKDVKASQPV--DADRQKQNYNT 2575
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 PMTST---IDLTKQNPVDKVNENHAPTYINTSPNKSIMKATPKASPKVAFTVTPET 179
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2576 AVYNAEIIINATSQ-PLDRBSATQANQVST--NKALNGAQNLAN-KQOETTANINQL 2631
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 HHPRDRVEED-----OQOQKEDSVPEPLIHOQWD--PSQFNYS 218
| : : : : : : : : : : : : : : : : : : : : : : : :
| 2632 SHL--NNAQKQDINTQVTPNPNISTVNOVYTKAEQLDQAMERLINGIDQDKQSVNFT 2689
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 DEDTNAVPTPLHTTKPRFAO-LINKNNEVSEPE---ALTDKLRKENSNTSLDE 273
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2690 DAD-----PEKOTAYNNAVTAAENITINQANGTANOSQVEALSTVTTTKQ--ALNGDR 2741
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 274 KYNLYLPTNNNNKSNVSDMSHLQLODASKNKTENHNLSPALKAPKNDIENPLSL 333
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2742 KV---TDKNNANQTLSTLD--NLNNAQKAVTGNL-----NQAHV 2778
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 334 TNDISLRSSGSSQSLSRNDNRVLESYPGSPKVNPLSLND---GIGFSDVV 388
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2779 AEVTAQIQTQAEIINTAMGNLKN-----SLNDKDTLTGSONFADA-- 2817
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 389 ESLLPRDLSDKLETTKEHDAPEHNNENFDKSTNTN--KGOL-----LVSS 434
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2818 -----DPEKKNAVNEAVHNAENILN-KSTGTNPVKDQVEAAMOVATKALNG 2865
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 435 DQHLDSFDSYNTTEOSILNLSASQS-----QISLMAEKQRTQEOEQTAAREEE 489
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2866 TONLEKAKGHANTAIDGLSHLTNAQKEALKQLVQGSTVAEAGNBOKANNVDAAMDCLR 2925
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 490 TSPSDNITVKOEKSNLEKVKYIKKEPVSAITKAPKEFSRIIRIKN-DEEIEPAD 548
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2926 QSLADNATTKQ--NONTTDASQNK-----KAVYNAVTTAGGIIIDOTTSP- 2969
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 549 IHPKKEANSHEVEDTALLKALKALNDEESDTQNTSKSIRPHIDSQKLEDSNDGRE 608
| : : : : : : : : : : : : : : : : : : : : : : : :
| 2970 LPTVYNQAGOVST---KANLNGENLEAKQAQASQSL-----GSLDLNNAQKQ 3018
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 609 D-NDDISREKSDILNDVQSTSDIIGDKYGNSSSEITTYTL-----APPRSDN 655
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3019 TVTDOINGAHATVDEANOIKONQONLNTAMGNLQALADKATKATVNTDADQAKQOAVN 3078
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 656 NDKENSKSLEDPAN-----NESLQOOLEVPHRKEDDSILANS--- 693
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3079 TAVTNAENIISKANGNATQAEVEQAIKQVNAKQALNNGANVQAHKKDATALINSNDL 3138
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 694 NIAPPELTLPVVEANDYSSFNVDYTKF---DAYSSFEESLSREHETDSKPIINFISIW 749
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3139 NQAOAKDALKQOVQNTTAVGVNVKQTQOELNNAMTQLKOGIADKQGT----- 3186
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 750 KQEKQKQKHQIHVPTKQIIASTYQOKKNEQESRVSQKVPNAIQKFKREVNVMSRRVY 809
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3187 ---KADGNFVNADPKQKNYVNAVAKAEALISATPDVVVTPSEI-----TALNRY 3234
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 810 SPDMDDLANSQFLPELSEDSGFKDLNFANYSNNTNPRSFPLST--KNVLSIINDPVP 867
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3235 TQAKNDLN-----GNTN-----LATAKOVVNAHIDQLPWL 3264
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 868 VEPPEKSYAEIRNARLSANKAAPNOAPPLPPOROPSTRSNSKRVSRFRVPEIR 927
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3265 NQAOQDEISKQITQA-----TLVPVNAIQQAATTLTDAMTQLKOGIANKAQITG 3314
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 928 TSSALAPCDMYNDI-FDEFGAGSKPTIKAGM---KTLPSMDKDVKRIILNKKGVTODE 983
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3315 SEN-----YHADDTDKQTAVDNAVTAKAEELLKQTTNTTMDPNTIQAQLTVNVDINQAL 3367
| : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 984 YINAKLVQK-----PKNSIVTDPREYRELOOTASINATIDSSITYGR 1028
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3368 NGNQKLDADAKQDKATLTGLTDLHNDQAQKALTT---QVQADRIATVNVKQNAQWL-- 3421
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 1029 PDSISTMPLVLSDELKPKPTALLSADRLFMEDQVNRPLRSNVLVHFGAATNSSMLPE 1088
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3422 -----NNAMTNLNNALQDKTETLNSINFETDADAKKDAVYTNNAVSHAEGILSKANGSNASQ 3476
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 1089 PDPE---LINSFARNVSNNSDNVAISGNAS-TISFNQDLNMFDDQATTQKTOEQPASK 1143
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3477 TEVEQAMORVNE-AKQALNGNDVNOQRKDAKQVITNANLQNAQKALQOVDAAQVTA 3535
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 1144 SANTVRGDDGLASAPETPTPK-KESISSKPAKLSSASPRKSPIKISPVAVIKKNGS 1202
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3536 NVMTIK-----QTAQDLNQAMTQLKOGIADKQO-----TANGN 3569
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 1203 IAGIEPRTKATIKPKKSFQSGNEISNKKVDRGDISPSSGSEHQNDPMSVPSQYTDATS 1262
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3570 FVNAD-----TDKQAVNNAVA-----HAEQII--S 3593
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 1263 TVPDEKNDQVKKPREKQK--QKHHNNHHNNHKKOTDIPGVNDELPRDVLQERKLFPR 1320
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3594 GTPNANVDRQVQAQALQVQAQKGLNGNHNLDQAKDMANTALDQLE----- 3640
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 1321 VLGIKNINLPDINTHGRET-LTLDNGVNCVTPPEVNMDDHVAIGKEBELTVADSLERI 1379
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3641 ---NLNPOKRTALKDQVSHAELVTGVNAI---KQNAIDLNNAMG----- 3678
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 1380 LTLK---ASYEKPRGTIVEVTEKKVYKSR-----NRLSRLEFGSDIITTTKVFVTEVK 1429
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3679 -TLKQOIOANSQVPOS--VDETQADODKQOAVYNNANQAOIANG-----IPTPV- 3725
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 1430 DTWANKFAPDGSFARCYIDLQCFEDQITG--KASQFDLNCFNEMETMSNGNOPKRGKPY 1487
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3726 -----LTPE-DTVTAQVNTTMMQAKDALNDEKELAQAQKQELANLDTLRDLNQPGRALNR 3778
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 1488 KIQOLEVKMLYVRSDPRELPTTSIRSAVESIETLNEONNFEQYGLHDE 1537
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 3779 QINQAOAL-----ATVEQTKOMQNVNTTAMSNILKOGIANKD 3814
| : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 2
US-09-815-242-5834
Sequence 5834, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308

APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 5703
LENGTH: 2025
TYPE: PR1
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(2025)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-5703

Query Match 3.5%; Score 298; DB 10; Length 2025;
Best Local Similarity 18.9%; Pred. No. 9,4e-08;
Matches 333; Conservative 246; Mismatches 663; Indels 524; Gaps 79;

17 LQLOPOSSASIFNSPKRP-----LNPRRTSKSLDP-----NSSSD 54
18 IODQOQTEAGSKFNEKPKDAYQAAVQAHAKDLNQTNPDLKAYEQLTGVGNQAKD 447
55 TYTSEODEKEKPKDTAFQTSFDRN-----FDLBNST---DIOQTIGHQOQRP-- 102
448 NLHDOQLADKKQAHVNDLQNLNSLNFQRALESQINNAATTRDEAVAKLEAQAQADAM 507
103 -----QQOQLSQTDNNLIDE-----FSFQTPMTSTLDLTKQ--NPTVDRV----- 140
508 QALNRSIQDQOQTESGSKFNEKPKDAYQAAVQAHAKDLNQTNPDLKAYEQLTGVGNQAKD 567
141 -----NENAPTYINTSPKNSIMKKATPKASPKKVAFTVTNPELHNYPD 184
568 VTTAKDNLHGDOQLARDOQAQVTTVNALPN---LNHAQOQALTDALINAPTRTEVAQHVC 624
185 NRVEE---DOSQOKEDSEVERPLQIHOHWKDPDSOFNYSDEDEINASVPRPRLHT---TKP 237
625 TATLIDNAHMETLKKKVDVN-----TKAQRPVTEASTDKKEAVDAQALQAEISTDP 676
238 TFAOLKKNNEVNSEPALTKMLRENFSMLIDEKVNLYLSPNNN-----NSKN 289
677 TNGSNANK-----DAVEQALT-----KIQEKVN-----ELNGERVAAEKAQAQKQ 716
290 VSDMDSHL-----ONLQDASK-----NKTENIHNLSALKAPKN----- 324
717 TIDOLAHNLMOAQTAKQONIDATKLPRIAEVLVDAQTLQNLQSMQDOLQAVVMEHNAVETQV 776
325 -----DIENPL-NSLTNADISLRSSGSSOSSLSLBN-----DNRVLESYGS 366
777 DYTQADSDKQAAQYAKALEAVNLKQNSNKQVQDQALQNLINAKQALNGDRAVALAKTNG 836
367 PKKVNPLSLN-----DGIKGFSDEVESLRLPRLDSRDLKLETTKEHDAPEHNNEFID-AK 421
837 KHDIDQALNALNNAQDQDKGRIDQ-----SHDL--NQIQQIVDQAK 875
422 STNTNKGQLLVSPDHLDSFDRSYNHTQISILNLSASQ--SQISLNALEKOROTQEQEQ 480

876 ALNRAMDQL-----SQEISNEGRTKST-----NRYNADTQKQVYDAVADKAKQALDKST 927
481 TQAAPEEETSFSDNIKQKQEKSKLFEVQYTKIKKEPSSAE-IKAPRRESSRLRKTN 539
928 GONLTAEQVIRKLN-----AVTAAKKALNGERLNNRSEALOR----- 967
540 EDEIAEPADIRPKKENEANSHEVDALLK--KALNDEESDPTQNSKMSIRFHIDSDW 597
968 -DQLTH---LNNAQQLALIQQINNAETLTKASRALNRTKIDNMG-----AVQOYIIDSQH 1019
598 KLESDNGDREDNDDISFEKSDILNDVSQTSIDIGDKYGNSSSEITTKTLAPPSDND 657
1020 LGVISTVYINADNKLKANYNAINAAHELDKVGQNAIAKAEAEQLQNIIDQALNG 1079
658 KENSKSLDPPAN-----NESIQOOLEVPHRKEDDSILANSNLTAPPELTPVYEADN- 710
1080 DONLNAKDKANAAEYNSLNGLNQOQODLH-----NAINMADTVSDVTYDINNQDILNDA 1134
711 -----YSSFNDVYTK--FDAYSFEESLSREHETSKPINFISTWH 749
1135 METLKLHVDNELPNAEQTVNQNMADNKTINFDQAKRLANTLINDNTNVNDIN----- 1188
750 KOEKOKKHQIKVPTKQIILASTQOYKNDQESRVTSDK-VKIPNAIQKKEKVEVNVMSRV 808
1189 -----GAIQAVKDAIQNLNGEORLOEAKDKAIOVNVKYLADKLKEIEASN--- 1233
809 VSPDMDDLNVSGFPELSEDSGFKDLNFANYSNNNRPSTPLSTKVVLSINDMDPVV 868
1234 -ATDQDKLIAKKAEELN-----SIINNTKATSNQDVQVQTAGN----- 1274
869 EPPEPKSYAEIRNARRLSANKAAPQAPPLPQ--RQPSSTRNSNKRVSREPVTF--- 923
1275 QAIQEVHANEIPKA-KIDANQDVQKQVQALIDEIGRPNLT--DKEQALDRINOIQQ 1331
924 -----EIRTSALAPCMYNDIDPDGFA--GSKPTTA----- 955
1332 GHNDINNALKREALEQAKERLA--QALODIKDLVAKEDAKNKIKALANAKRQDQINSND 1389
956 -----EGMKTLPMSMDKQVIRLNAKKGVTODEY---INAKL-----VDQKPKNS 998
1390 LTPBEKAKALEIDEAEKRALQNVENNAQTIDQNLNGILGDDIDNTNHWVEYDDQPAVNE 1449
999 IYTPEDREYEELOQTASIHNAITDSISYGRPSIST--DMLEPYL-----SDELKRPPT 1049
1450 I-----SEATPEQQLLVNGLIELVHRDITTEQDVLAHINLIDQLAEVDPST 1497
1050 ALLSADRLEFMEQVHPLNSNSYLH--PGAGAATNSMILPERDFELINSPAR---NVSNNS 1105
1498 ATIS-DSLTARVEVTLIDGSKVIVPVKVVYKELSVVQQAIESIENAVOOKINEINS 1556
1106 DNVAISGNASTIS-FNOL-----DNFPDQATIGOKIQOPASKSA- 1145
1557 VTLTIEQKEAALAEVKNKQQAIDIHINAPDVHSAVEEIQOEOAHIEQFPNEQFTIEQAK 1616
1146 -NTVNGDDGLASABETPRTN-----KESISRPKALSASPR----- 1184
1617 SNAIKSIDALQIHMDIKARTIDTFKEKEQAL--AKNLQKQALQIOLQRAQSIDEI 1672
1185 -----KSPIKIGSP-----VRYIK-----KNSIAGIEPIPKATKPKKSF 1220
1673 TEOLFQFAOKMAANPTAKELAKRQEAISIKIDFSNEKMSIRNSE-IGRADEQQAAMN 1731
1221 QGNELSNHKVRD-----GGI-----SPSSGSEHQHNS 1249
1732 QINELVLETTIRIDINNAPHLQOQVEALNNGIARISAVQITVSDRAKQSSSTGESHSH--- 1788
1250 MVSVPQYTDATSTVDEKNQVONKPREKOKNHNHNHNHNKOKTPIPVVDEIDPQV 1309
1789 ---LTIGV--GPAHPFNSSTIGHKKIDEDDDIDPCLMHRF---SNFGVNIKALICV 1840
1310 GLOERKGLFFRYVLGK-----NINLPDINTHKGRTLLDNGVH---CVTTPPEYKMD 1358

Oy 1146 -MTVAGDDGLASAPETPTPT-----KESISSKPAKLSSASPR----- 1184
Db 2748 SNAKSIEDAIOHMEIDEIKARTDLTKKEQEA1---AKLNQKQAOIOLQRAQSIDEL 2803
Oy 1185 -----KSPKIGSP-----VRV1K-----KNGSAGIEPIPKATHKKKSF 1220
Db 2804 TQEOLOFKQKMAANFTAKELAKRKQEA1SKIKDFSEKMNKNSIRNSE-IGTADKQKMAAN 2862
Oy 1221 QGNEISNHHVRD-----GG1-----SPSSGSEHOQHNPS 1249
Db 2863 QINEIYLETIRIINNANHILQOYEALANNIARISAVOITSDBRAKSSSTGMSNSH---- 2919
Oy 1250 MVSVPQYTDATSTVPDENKDVQHKPREKOKNNHHHHHHHKKOKTIDIPGVVDEIPDV 1309
Db 2920 ---LITGY---GTANHPFNSTIGKHKKKLDEDDIDPLHMRHF---SNFGNV1KNAIGVY 2971
Oy 1310 GLOEKGKLFERYLG1K-----NINLPDINTHKRFLTLDNGVH---CYTTPPYNMD 1358
Db 2972 GISGLASFWFIARKKRREDEEELEIRDN--KOSIKETLDDTKHPLFAKRRKKED 3029
Oy 1359 DHNVAIGKEFFELTVADSL-----FILTKASYEKPRGTLV--EYTEKKVYSRRRLSR 1410
Db 3030 EEDVYVEEKOSINTGESLDKVAKHTPFLPKRRKREDEEDVYTNENTDEKVLKDNHSPV 3089
Oy 1411 LFGRK-----DITTTKFPTEVKD 1430
Db 3090 LIAKRRKKEDEVEETTSI---ESKD 3112

RESULT 5

US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12713
; LENGTH: 2344
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match 3.4%; Score 289.5; DB 10; Length 2344;
Best Local Similarity 17.8%; Pred. No. 3.2e-07;
Matches 259; Conservative 247; Mismatches 708; Indels 237; Gaps 43;

Oy 18 QLOPOSSASIFNSPTKPLNFPRTNSKPLDPNSSDTY-----TSEDOEK 64
Db 951 KLESQSTSLSTSDSKSM-----STSESLSDTSTSDSVSGSLSVAGSOSVSTSDSMS 1006
Oy 65 GKBEKKOT-----AFQSTDRNFDLNSDIDIOQTIOHQOQOPOQOQOOLSDTNNLIDFS 119
Db 1007 TSEMLSDSMSTSGSLAASOKSMVSSSMTSGOS--GSTESLSDS1STSDSKSLST 1065
Oy 120 FOTPMSTLDTLRKONTYVKNVNNAPTYINTSPKSIKKKATPKKPVFTYVNPET 179
Db 1066 SOSGSTSTSTSSVYRMSESOSTSGSMSTSGSDSTISSTSDSTSDSKASTASSSEI 1125
Oy 180 HHYPDNRVEEDQSOQKEDSVPEPLIQHOKDPQOPNVEDPTNAPPTPLHT--TKP 237
Db 1126 -----SOSVSTSG-----TSGSVSTSLSTSNSR 1150
Oy 238 TFAQLLNKNNVNSPEALITDMKLRNFSN--LSLDEKYNLISPTNNNSKNVSDMDSH 296
Db 1151 TSTMSDSTSLSTSESDSTSDSTSDSTSDSISEA1SGSESTSLSESNSTSDSESKASAF 1210
Oy 297 L-ONLQDASKNKTNNIHL-----SFLAKAPKNDIENPLSLN-----ADISLRSGSQ 347
Db 1211 LBSLSLESTSESTSESLSGSTSDSTSLSDSNSESGSTSTSLNSTSGASISTSTSGAS 1270
Oy 348 SSU-----QSLRNDNRVLESYPGS--PKRVNQLSLNDG1KGFSDY 387
Db 1271 TSVYKSESVSTSLSTSTSLSDSTSLSTSLSDSTSGASNSLSAMSTSDSTSTKSES 1330
Oy 388 VESLLEPRDLRDKLETTKHDAPENHNEFIDAKSTNTKKQLVSSDHL--DSDRSTN 446
Db 1331 LSASTSLSGSTSESGSTSSSESKSDSTSMLSMSQSTSGSTSVSTSLSDSTSTSL 1390
Oy 447 ---HTFQSTLNLSASOS---QISLNALEKQRTQEQTOAAPEEETSPDN1KVKQ 500
Db 1391 LSAAMNQGVD--SNASQASASTSTSTSESQSTSSSTSTSTSTSTSTSTSTSTST 1449
Oy 501 EPRSNLEFVKVYTIKKEPVATEIKAPKREFSSRIIRKNEDEIAPAD1HPKREANSH 560
Db 1450 S1KSTSGSGSTSTSA1SLSGSESDSQS1STSESKSESTSTSLSD---STSTSNHG 1505
Oy 561 VEDTDLALKALNDDEESTONSTKMRFRHIDSMKLEDSNDGDRENDOD1PREKSD 620
Db 1506 SASSTSTLS---NSASASSDSSSTLS-----DSTASMSQSESDSTSTST 1550
Oy 621 IINDVQSTDI--IGDKYGNSSSEITTKTAPRSDNNDKNSLDEDDPANNSSLOQLV 679
Db 1551 LNSQSTSTSIKMTIASVSESESTSESGSTSESTSESTSTSLD---SQSTSRSTA 1607
Oy 680 -----PHTKEDDSILANSSNIAPPEELTP---VEANDYSSFNDVTKTFAYS---- 725
Db 1608 SCSASTSTSDSRSTASTSTSMRTSLDSQMSLSTSTSVSDSTSLSDSVSDSTSD 1667
Oy 726 -----SFEESLSRHEHDSKPINFISIMHKOEKKKQIHKVPIKOLIAS1QOKK 775
Db 1668 STSTSTSGMSASISLSDSTSTSTSAEYVMSASISDSQMSSESVNDESVSSNESDSK 1727
Oy 776 NQOESRVTSDKYKIPNAIOFKKFEVY-----VMSRRVYSPMDLANVS--QFLP 823
Db 1728 SMSGSTSVSDSGSLSVSTLKRKSEVSEISLSGSGSMKSDSVSTSDSSLSVSTGORSSE 1787
Oy 824 ELSEDSGFRDLNFANYSNNTNRPSTPLSTKYNVLS--NIDNDPVNVEPPEPKSYAEIR 880
Db 1788 SVSESDSLSDSKSTSGSTSTSGS---LSTSTSLSGSESVSESSLSDSISMSDSTST 1844
Oy 881 NARRLS-----ANKAAPNAPPLPQORQPSSTRSNKNSKRVPRPTFIRRTSSALAC 935
Db 1845 DSDSLSGSISLGSSTSLSDSLSDSKSLSSQSMGSESTSTSVSDSSSTSTSTST 1898
Oy 936 DMYNDFDDFGAGSKPTTIAEGMKTLPSMD-----KDOVKRIILAKKGVTD 982
Db 1899 ---NSOFDSMSISAS---ESDSKSTSDSSISGNSSTSTSLSTSDSMGSVSVSTSTSL 1952

QY 983 EXINAKLVDOEKPKNSIVTPEDRREELQOTASIHNAITDSSI-----YGRPDS 1031
 DB 1953 DSISSISIVSDSSSTSTSSLSOSMAQOSOSTSTSAAGSLSTSLISMSASACTLTQSST 2012
 QY 1032 ISTDMIPYLDELKPPYALLSADRLPMEQEV---HPLNSVYLHPGGAATNSSMLPE 1088
 DB 2013 VSTSL--STSDSISTSTISISGSQSAVESESTSDSTSIDSESLSTSGSTSSSTSTST 2070
 QY 1089 POFELINPARNV-----NSDNVALISGNASTISPNQIDMNDQATIG 1133
 DB 2071 EELSTISNCSSTSVSESLSTSGSGSTSVSDSSSTSSLSSTSGSTSVSDSTSMSESNASIS 2130
 QY 1134 OKIOEPASRANTVYRGDDGLASAPETPTP--TKESISSKPAKLSSAPKPSIKIG 1191
 DB 2131 MS-QSISGSGSTSTSTSESLSMGSGTINSTSVSDSDSTSTSGSGSMNSIHFTSLST 2189
 QY 1192 SEVRRIKKNGSIAGLEPIPKATHKPKKSPQGNISNKKYRDCGISPS---SGSEHOQHP 1248
 DB 2190 SGLMSLSNMSMSDSVSISASESMSASMSDSVMSDSTSSMSMSKSTSESNISHP 2249
 QY 1249 SNVSVSQVTDATSTVPDENKVQNHKPREKOKHHHHHHHKKQTDIPGVVDEIDP 1308
 DB 2250 SN-SMSTQHTSTSTSTSTSESTAPNTNESOSTLSATSVSKHDAE---PAOSEERLDP 2304
 QY 1309 VG--LQERGKL 1317
 DB 2305 TGDSTIKONGL 2315

RESULT 6

US-09-820-843A-15
 ; Sequence 15, Application US/09820843A
 ; Publication No. US20030039963A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Council of Scientific and Industrial Research
 ; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
 ; FILE REFERENCE: Q63915
 ; CURRENT APPLICATION NUMBER: US/09/820,843A
 ; CURRENT FILING DATE: 2001-03-30
 ; NUMBER OF SEQ ID NOS: 118
 ; SOFTWARE: Patent version 3.0
 ; SEQ ID NO 15
 ; LENGTH: 1139
 ; TYPE: PRT
 ; ORGANISM: M. genitalium
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: cycladherence-accessory protein
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: g111046012
 US-09-820-843A-15

Query Match 3.2%; Score 280; DB 9; Length 1139;
 Best Local Similarity 19.0%; Pred. No. 4e-07;
 Matches 225; Conservative 188; Mismatches 435; Indels 334; Gaps 53;

QY 39 PRTNSKPSLDPPSSSDTYTSEDOEGKEKKTAFOTSDFRFDLNSI--DIQOTIGH 96
 DB 184 PEVKQESVVDQSSDDYFAKOPT-----DENYGFNDNDLPEVKOP-ES 226
 QY 97 OOOQPOOQOQLSOTDNLLIDEFSGFQTPMTS--TLDLTKONPTVD---KVNENHAPTYINTS 152
 DB 227 VVDQPSDDHFAKQPESTSDSYSFSDSLDLPQPLTD-----QPSLDHNYQYNHDI----- 274
 QY 153 PKSTIMKKAATPKASPKKVAFTV-----TNPETIHNYPDNKEVEEDOSQOKE--DSVEPPL 204
 DB 275 --HEELKPYAEERQNNYQVGFDOVQANLDMNEETIQPAEKKVTTDPEFSKQAVVDYOLPI 332
 QY 205 IOHOMKDPQOFNYSDDEDTNAVSPTPLRTKTKTFAQLLNKNNEVSE--PEALITMKKL- 261
 DB 333 VTDQ--ODQTFSSSFE-----TQPTVEQFDQVNSEVNDQEPKELTKPEVLE 377

QY 262 ----KRENFSLDEKVNLYLSPFNNNNSKAVSDMDSHLQNTODASKNKT-----N 309
 DB 378 SSFNKQDVYETSDLNSESNIY--SENNKQATNNDSLNEFTQLNNSNETASDYHYEKS 435
 QY 310 ENIHNLSPALKAPKNDIENPLNSLTJNADISLRSSGSSQSLQSLRNDNRYLESVPGSPKK 369
 DB 436 EPIHDYKF-----GSDLSQS-----NSNNLES---EPVK 462
 QY 370 VNPGLSLNDICKQFS---DEVESLCPRLDSRKLKLTTHKD---APEHNNEFDIAKS 422
 DB 463 FNSETAPDAHFESQSEPVDOYQYDIQNELKPTLPDQSSDDYFAKQPTDENGFQNDLP 522
 QY 423 TNNKCOLIV---SSPDHL---DSFGRSYNHEQSTILNLSASQSLNLAEKORQT 475
 DB 523 PEVKQESVVDQSSDDHFAKQPESTSDSYSFSD-----LPQPLDQPSLDHNYQYNFD 577
 QY 476 QEOQOTQAAPEBETSFSDNIKVKQEPKSLFEVYKTIKKEPVASATEIKAPKEFSRIL 535
 DB 578 HHEELKPYA--EERNNYQVGFDOVQANLDMNEETIQPAEKKVTT-----DFESKQA 626
 QY 536 RIKNEDEIAPADIHKKEANEANSHEDTALLKALNDEESDITQNSTKMSIRHIDS 595
 DB 627 QVVDYQYL--PIPTDQDQDTTFSSSFEPTQTV-----EOPDYVNSEVNDQPKPEITK 676
 QY 596 DMKLEDSNDGDRDNDISFEKSDILNDVQSTDIIG--DKYNSSEITTKTLAP----- 650
 DB 677 EPLLESS-----FNKQDVYETSNYNNLQKPIQSDNKITITTKKSSQIPT 723
 QY 651 --PRSDNDEKNSKSLDDPA--NNESTLQQLLEVPHRKEDDSILANSNIAPEELTLPV 706
 DB 724 TLPISFVSNRIEYKPVETLALDKKESQEOITINSTIEDSKTLAKT-----LSVQL 775
 QY 707 EANDYSFNDVYKTPFAYSSFEESLSREHETDSKPINFISIMKQEKQKHQHKVPTQ 766
 DB 776 QINSLNQSTIVT-----SESVRLDKKDDQLTINTV--NSEDQAPRIEYV--YKAE 822
 QY 767 IIASYOQYKNEQESRYTSQVKIPNAIQFKK---FKEVWMSRRVSPMDLNLVSQFL 822
 DB 823 PVEHSTQTKQKQ---VEDKSELDN---FNKKGDLYIISELRGELINPINFDAIQMN 876
 QY 823 P-ELSEDSGFKDLN--FANYSNNTNRRPSFTPLSTKNVLSINDNPVPEPKSYAET 879
 DB 877 DYQMSVQSEFIHLNDEVTNKNQISE----- 902
 QY 880 RNARLSANKAARNQAPPLPQORQPSSTRSNSKRVSRFPVFEIRRTSALAPCDMYN 939
 DB 903 ---RYLLIKKELOSELRLIDQNEMLNVQFNNAKNL-----TTLQKEEMIR 945
 QY 940 DIFDDFGAGSKPTIKAGMKMTLPSMDKDVKRLINAKKGYTOBEYINAKLVDOEKPKKNSI 999
 DB 946 SLASDFALAKPNSNSYEQQL---KSGELMR--HYQRAITENE-----KR--- 984
 QY 1000 VDPEDRYEELQOTASTIHNAITDSSIIYGRPDSISTDMLPYLSDELKPPYALLSADRLFM 1059
 DB 985 ---IESIOGSLKOLKTYVYNSCCE-----IMNNINKLIDTLL-----RFA 1020
 QY 1060 EOEYVHPLR-----SNSVYLHPGGAATNSSMLPEPDELINSRANVSNNSDNVAISG 1112
 DB 1021 KKEKDPILLSNFDVYDNGLVEP-----NQLMDLIDL-----SNTFDNIS-NE 1063
 QY 1113 NASTISFNQIDMNF-----DQATIGQIKQEPASKSN 1146
 DB 1064 QLDQFTYENMDRNIDPEFEGFNNDPVIDAKVMDSKASFSVN 1105

RESULT 7

US-09-815-242-5635
 ; Sequence 5635, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.

RESULT 10
US-09-764-176-7
Sequence 7, Application US/09764176
Patent No. US2002012753A1
GENERAL INFORMATION:
APPLICANT: NOTEBOURN, Mathieu Hubertus Maria
APPLICANT: DANEN-VAN OORSCHOT, Astrid Adriana Anna Maria
APPLICANT: ROHN, Jennifer Leigh
APPLICANT: WEISS, Bertiam
TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
FILE REFERENCE: 472505
CURRENT APPLICATION NUMBER: US/09/764,176
CURRENT FILING DATE: 2001-01-17
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.0
SEQ ID NO 7
LENGTH: 1400
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Description of Sequence: Amino acid sequence deduced from the nuc
HS-09-764-176-7

Query Match 3.2% Score 276.5; DB 10; Length 1400;
Best Local Similarity 19.4% Pred. No. 8e-07;
Matches 275; Conservative 226; Mismatches 514; Indels 403; Gaps 72;

QY 41 TNSKSLDPNSSDYTSODEKGEKKOTAFQTSFDRNFDLDN---SIDIOQTIOH 96
DB 183 SRESSESL-----DEETKKEETPKQOEKSEKMSQEPMDLEKRSANVALEETVAK 237
QY 97 QQQQPOQQOQSLQDNNLIDFESFQTPMTSLDLTKONPTVKNVNAPTYI----- 149
DB 238 EKEDEKELVLRP-----VIVKLEKPLPENEKKIKKEES--DSFKENVAPKIVEVEKCA 290
QY 150 NTSPKSTIMKKTTPKASPKKVAFTYNPILHYPNKRVEEEOQOEKESVEPPLIQHOM 209
DB 291 DKDKTKSSMEKPVAG--EPRRIEFG--GNIKSHETEKSTEEKELKNDQAAKIPKKREI 348
QY 210 KDPQSFN-----YSDDTNASVPPTPL--HTTKPTFAOLLKN 246
DB 349 KLSDFDPSVKGRLCKSVTPTEKFLKDEIKOEETCKRISTITALLGHEK-----QIV--N 402
QY 247 NEVNSE---PEALTD-----MKLKRENFSLDEKVNLYLSPITNNNSKNVSDMSHL 297
DB 403 GEVSDEKVAPEKTEPIETKFEYETKEESYSP--SKDRNITTEGNGTESLNSVITSMKTGEL 461
QY 298 ON-----LQDASKNKTNENIHLNLSFALKAP--KNDIENPLNSLTNA--DILRRSSGSSQSL 350
DB 462 EKETAPLRKRDADSSISVLEIHSQAKOIEEDPPEMETSIDSSSEMAKDLSKSTALSSTESC 521
QY 351 QSLRDNDFVLESVPSPKVPKPNGL-----SLNDGIKGFSEVEVSL--LPRLDSRDKLET 403
DB 522 -TMKE-----EKSPKTKKDKRPPIECLEKLEKSKKTFIDKQARSLRPIPEVPKLTLS 576
QY 404 TK-----EHDAPEHNNEFIDAKSTNTKGGQLLVSSDDHLDFDRSYNH 447
DB 577 EKPGSPKAEETSPSPNIIDHCEKLEKSEKVEVCEQSTJVGQGSYKVD--LETLKEDSEF 634
QY 448 TEGSLNLLNSASQOISLNALEKORQOEQOTQAAE-----PEERTSSDNIKYQOEK 503
DB 635 TKVEDNDLNDQOT-----SGLEEPSKTKGSMQSKFKKVLVPEETJASSENTTITSEK 688
QY 504 SNLEFVKYTI-----KKEPVASATEI-----KAPKREFSRILR--IKNEDEIAP-- 546
DB 689 K--EGIKLITIRSSKKKRPDPSPKVLPEPNKOEKTEKEEKINVGTLIRSRISPTAK 746
QY 547 ADIH-----PKKENANSHVEDTDALL-----KKAINDDESDTTONSTKM-----SIRF-- 591

DB 747 VAEIRDQKADKKRGEDEVEEESTALQKTDKKELKKSEKDTNSKVYKPKGVYRWTG 806
QY 592 -HIDSDMKLEDSNDGREDNDDISREFKSDILNDVQSTSD-----IIGD-----KIG--- 637
DB 807 SRTGRMKYSSN-----DESEGSSEKSSAASEEKEKSEBALLADDEPKCKGCLPN 860
QY 638 -----NSSEITTKTLAPPSRDNNDKE-----NSKSLDPPANNELOOLEVPH 682
DB 861 HPELIIILCSDSCSYHTACLRPLMIIPQGEHMCPCPOKHLLEKTL--EQQLDLDVALK 918
QY 683 KEDDS-----ILANSSNIAPPELTLP-----VVEANDYSSEFNDV 717
DB 919 KKEARERKERLYVYGISEIENIIPQEPDSEDOEEKKSKKANLLERSTRKCI 978
QY 718 TKTFDVASSSEESLSREHEDSKPINFISIMHQEO--KKHDIHKVPTKQIIASIQOYK 775
DB 979 SYRFD--EFDEAIDEAIEDIDIKADGGGGRGKIDISTITGHGKIDIST--ILDERKE 1032
QY 776 NEQESVTSQKIPNAIQPKFEVNVMSRVVSPMD--DLNVSOFLPELSED-----S 829
DB 1033 NKRPQRAA-----AARKKRRRLN-----DLSDSNLDE--ESEDDEKISD 1072
QY 830 GFKDLNFANYSNNTNRPSEFTPLSTRKNVLSNIDNPNVVEPPPKSYAEIRNARLSANK 889
DB 1073 GSQD--EFVVDENPDSEEDPD-----SNDSQDTDFC-----SRRLRRHP 1111
QY 890 AAPNOAPPLPPOPOPSSTNSNKRVRFRVPFEETIRRTSSALAPCDMDINDIDDFGAGS 949
DB 1112 SRPMRSRRLRRTPPKKYSDDDE-----ESEEENSRDESDFDSDDFD-- 1158
QY 950 KPTIKAEKQTLPSMDKDVKRLINAKKGYOTODEYINAKLVQDKPKKNSIVTPPEDRYEE 1009
DB 1159 -----VETRRRRRRNKKROINKKEDESG-----SOKSLR-----RGKE 1194
QY 1010 LQQTASIHNAITDSSIYGRDPSISTDMLPYLSDELKPPALTALSADRLFMEQEHVPLRSN 1069
DB 1195 IRR--VHKRRILSS-----ESEESTLSKN 1216
QY 1070 SVLVHAGACAAITNSMLPEPDFELINSPARVNSNDNAISGNASTISINQIDMNFDDQ 1129
DB 1217 S-----EDDELAESKRSYKR-----GRSTDEYSADDEEEEE 1250
QY 1130 ATIGQKIQOPASKSANTVRGDD-----GLASAP--ETPRT--PTKSESISSPAKL 1178
DB 1251 -----EGKPSKRLRHRIETDEEESCDNAHGANGAPARDSQRVLPSEDES--TKKPYRI 1302
QY 1179 SSASPR--KSPIKIGSP-----VRVIKNGSIAIGIEPIPKATHPKKRSFQ--GNEISNH 1228
DB 1303 ESEDEEDFENVGKVGSPPLDYSILVDLPSTNGQSPG--KAIEHLICKPTEKSQTPKDNSTASA 1361
QY 1229 KVRDGGISPSGSEHQHNPMSWYSPQYTDATSTYPD 1266
DB 1362 SLASNG--TSGQGE-----AGAPEEDEDLRLVTD 1389

RESULT 11
US-09-820-843A-16
Sequence 16, Application US/09820843A
Publication No. US20030039963A1
GENERAL INFORMATION:
APPLICANT: Council of Scientific and Industrial Research
TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PRO
FILE REFERENCE: 063915
CURRENT APPLICATION NUMBER: US/09/820,843A
CURRENT FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 118
SOFTWARE: PatentIn version 3.0
SEQ ID NO 16
LENGTH: 1616
TYPE: PRT
ORGANISM: M. genitalium
FEATURE:

NAME/KEY: misc_feature
OTHER INFORMATION: cytoadherence-accessory protein
NAME/KEY: misc_feature
OTHER INFORMATION: g11046097
US-09-820-843A-16

Query Match 3.18; Score 271; DB 9; Length 1616;
Best Local Similarity 19.08; Pred. No. 1.9e-06;
Matches 287; Conservative 259; Mismatches 569; Indels 396; Gaps 78;

39 PRT--NSKPSLDPNSSSDTYSSEDOQEK-GKEKKDFAFOSFDRNF--DLDSIDIDIOQT 93
1 2 PRTTKNNKNTTPKSKTKKYLESAKSKSVTPKKEQDQVEMLPDQPLIGETKKNT-LKKT 60
94 IQHQOQOPOOQOOLQSDNNLIDEFSPQTPMTSTLDTKQK--PTVDKVN--ENHAP 146
61 KSFNSKKKEVYKSKSPIDFFDE-----TKRGVFLVPTETDLSRELNQK 107
147 TYTTSRKSIMKATPKASPKKVAFTYTNPEIHHPNRYEEDQSQKEDSVPEPLIQ 206
108 TVVTVPNQT--SSYPTINENKLV-----ELNNQPETKYLE--TKKDSFTTITRE 153
207 HQ-----WKDPS-----QENYSDQEDTNASVPTPLHT 234
154 KKLMPDSQAWYIFVGDRAKYGPKNHTWVLGYFDOLQKNNKVEVITVEVPOE--HA 210
235 TKPTFAQLLNKNEVNSPEALTDMLKRENFSLN-----LDEKVNLYLS 280
211 A-----FKQRP--ADIDFWRPVGNPNNGYGVONNTWIMKGFDDKLNWIPI 255
281 PTNNNSKNSVSDMSHLQNDODASKNKTENENIHNSFLKAPKNDIEPLMSLTADISL 340
256 PV-----RTEELAGHDSLVDELEKKT-----ISEQPYWQENDIIVTVFNTKSL 301
341 RSSSSQSSLOSLRNDKRVLESVPSPKV--NPGSLNGIKGFSQEVESLLP----- 393
302 ASSLENELLENSSEQGVIEEVKPRNVEYIFRNVTKLH-----FEKEKEFLNPVRET 356
394 -----RDLSDKLETTKEHAP--EHNNEFTDAKSTNTKQGLVSSD--HLDSEFDS 444
357 NETTIPLEIVEEYKVESEVAPETPEACEPETTIPRETVGVYVEDDLKGLDS---- 412
445 YNHTQOSTLNLNSASQISLNLLEKOROTOEQOQTAAREEETSTSDNITKVKQEPKS 504
413 -NQTOAGNPEVEYFVEDDLKGLDSITIK--DDQOHDEIAHVEHLSODYSKEIKDSAKA 470
505 NLEFVKVTKI--KEVSAPEIKAPREFSSRIIRIKNEDEIAPADHPKKEANA--N 558
471 DLSNIDSDIDSVKKEFGSFTDETOKSVESKOV-----DELT--LDANNDFINESLRD 522
559 SHVEDTDLALKALNDE--ESDTQNSTKMSIRPHIDSMKLEDSNDGDRDNDISRFE 617
523 EVVNNIDQINIEYSEQGFEPYSVNEFOQEFSEPVVSDEKIKETN--SDSVNTDLALF 581
618 KSDLNDVQSOTDIIGDKYGNSSSEITTKIAP--PRSD-----NDKENSLSLE--- 665
582 SEKLNVNELITNEVY--DVNAPESTETEYKVSSELPKSELVDEITFINNDPKPOELEKXV 640
666 -----DPAN-----NESLQOOLEVPHTKEDDSILANSSN----- 694
641 DFLTEPKSLDEKTTIYVESEPPPIQDLSLEDSVNDVDSLETKTTSVELNHEETGN 700
695 -----IAPPELTLPVVEANDYSSFNVDTKTFDAYSSFE-----ESL 731
701 EFINLDVSEKVEQOPPTQLETFDSEFVLPYQIVE--DSFTSAELRPNFFSSQOKTLEFI 759
732 SREHETDSKPIINFISIMKOKOKKHQIHK-----VPTQIITASIQOQYINEDSRYTS 784
760 SQOTEVETSESNVPTV--EOETKLFEPHODENNLFTPLPLDLEITIEISNMLFSPKPEKSS 818
785 DKVAPNAIOKKKPKF-----EVNVMRSRV-----VSPD--MDOLNVSQFPELSEDEGF 831
819 DSELOPT--FKETKLDSTVEVPOESSOVEATFTVDQPEAVFDELKIOELOPEATTEVVF 875

832 KDLNFANYSNNNTNRRSFTPLSTRKVLNSINDNPNVPEPPEKSAEIRNARRLSANKAA 891
876 DD-----HFQPDVQPEQ--TPQEAKE-----FDSF--VEIPQESSQAEF--HMEQISDELKL 920
892 PNQAPPLPPOQOPSSTRNSKNKRSRFRVPTFERIRSSALAPCDMNDIDFDGAGSKP 951
921 EKETEAVDHOOLEN--QSEETVTPPEVTAFFETETLEQLEP----- 961
952 TIKAGCMKTLPSMDK-----DVKRIKLNK-----GVTOEYINAKLVQDKPKKNSIV 1000
962 --SSQDQSEPALQONNHEIYTAEEQIFDGTKLEDLKLECANDNENNEVQKRE----- 1015
1001 TDPE--DRYELQOTASIR-----NATIDS-----SIYGRDSISTDMLPYLS 1041
1016 TEAETFDKETQEOETSEPELSTEELKSEATFDVNSAEAEVPEK--QETQETEKILE 1074
1042 DELKRPFTALLSADRLFMEQVNHPLKRSVLYNHGAGATSSMLRPFDFELINSPARNV 1101
1075 EEPKSEP-----VQDLTEASFQTVKHEAVF--DKNQOTEGLEEQ--VSSEAEV 1122
1102 SNSNDVAISGNASTISFNQ--LDNMFDD--QATIGQ--KIOEOPA-----SK 1143
1123 DQTTIDIV--GEPEAVFPVQPEKTEVAFEDVENQKAYISEPVQEQGEAVFERSAEAK 1180
1144 SANTVRGDDGLASAP--ETPTPTKESISSKP-AKLSSASPRKSP--IKGSPVRY 1196
1181 FDSPEVSVQD--SQPEFVLEEVQTOPEIQVESQPEATFTVDQPEQKAFKPSVET 1237
1197 IKKNSINGIPIPKATPKPKKSFQNGEIS--NHKKRQDGI--SPSSGSEHQNHPS 1249
1238 VEQ-----PEFSEPTQOHVESASFDPEPVYDFDENYDFDQDSYSDIQPSEPO 1287
1250 M-VSVPSQYDTATSTVPDENKDVQHKPREKO-----KQNHNNHNNHNNKQTDIP--GVV 1302
1288 YDQDEPNVDFO-----EPNVEISKSEPOFEPVQEQPEAVFERSAEAKFDSPEVSV 1341
1303 DDELIPDVGLOE 1313
1342 QDSQPEPLLEE 1352

RESULT 12
US-09-987-482-1
Sequence 1, Application US/09987482
Publication No. US20020184656A1
GENERAL INFORMATION:
APPLICANT: BHANDART, POONAM
APPLICANT: SHAHIDHARA, L.S.
TITLE OF INVENTION: IN VIVO ASSAY SYSTEM FOR SCREENING AND VALIDATION OF
FILE REFERENCE: 056859-0134
CURRENT APPLICATION NUMBER: US/09/987,482
CURRENT FILING DATE: 2002-03-21
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 2843
TYPE: PRT
ORGANISM: Homo sapiens
US-09-987-482-1

Query Match 3.18; Score 269; DB 9; Length 2843;
Best Local Similarity 19.18; Pred. No. 4.8e-06;
Matches 287; Conservative 215; Mismatches 571; Indels 428; Gaps 64;

22 QSSASATNSP--TKPLNPRNTSKP-----SLDPNSSSDTYSSEDOQEK 66
924 RSSAAHTSNNTYNTKSSNSNRTCSMPYAKLEYKSSNDSLSNVSSDGY--GKRQMKPS 982
67 EKKQTAFOTSF-----DRNFDLNSIDIOQTIQHQ-----QQ 100
983 IESYSEDESKFCSTYGOYPADLAKHTISANMDDQGLDPTPINTSLKYSDEQLNSGRQS 1042

```

QY 101 P00000LST0NNLIDEEFQTPMTSLDTLRKONPTVDKVENNHA-PTYINTSPKRSIMK 159
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1043 PS0MERNAKRPKHIIIEDEL-----K0B0R0SRN0STITYPYTTESTDDKHL-- 1087
QY 160 KATPKASPKKAVFTVNBEPHIIHYPDNRV-----EED0S0QKEDSVE---PPLI0H0W 209
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1088 KFOPIFG0Q0ECVSPYRSGANGSETRVNGSNNGINQNV0S0LC0QDDYEDDKPTVYSEXY 1147
QY 210 KDP0QFNYSDEDTNAPSPTPLPTTKPTFAOLNKNNEVNSPEPALDMKIKRENFSL 269
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1148 SEED0HEEREPYNSI-----KYNEEKRHVQ0PIDYSLKVA--TDI 1187
QY 270 SLDEKVNLYLPTNNNNKNSVSDMDSL0NL0DASKNKTNN--IHLSPALKA--PKNDI 326
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1188 PSS0K0SFSSKSSG0SKTEHMSSEENSTPSSNAKR0NL0PSSA0SHSG0P0KAA 1247
QY 327 ENPLNSLTNADISLR-----SSGSSQSL0SLRND--NRVLESVP0GPKRVNPG 374
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1248 TCKVSSINQETIQYCYVEDPTICFRCSSLSLSAEDIGCQNTQ0EADSANTLQIAEI 1307
QY 375 SLN0GICFSD0VE---SLPRLSDKLETTKEHDAPEHNNENFID-AKSTYNNKQ 429
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1308 KEKIGTRSAEDPVSEVPASQHPRTKSSRL0GSSLSSESARKAVEFSSGAKSPSGA0 1367
QY 430 LLVSSDDH-----LDSEFDRSYNHTEOSILNLNSASQ0-----ISLN 467
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1368 TPKSPREHYVETPLMSEKRTSVSSLDSE-----SRSIASSV0SP0SGVSG0IISPS 1421
QY 468 AL-EK0R0T0E0T0A0A0PEE0TSPFSDNIKVK0PKSNLEFVKYTIKKEPVATEIKAP 526
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1422 DLPSPG0T0P0PSR0K0T0P0P0TA-----QTKREVPKN-----KAPTAERESGP 1467
QY 527 KR-EFSSILIK-----NEDEIAPADLHPKKE 554
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1468 K0AAVNAV0QV0VL0PADTLHFAT0ESTPDGFCSSSL0SL0DEPFL0QKVELEIRIMP 1527
QY 555 NEANSHVEDAL0K0KALNDEESDT0N0T0KMSIRFHD0M0KLEDSNDG0REND0IS 614
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1528 VQ0ENDNG0ET0E0PK0EN0E0K0A0E0KT-----IDSEK0LDD0S-----DD0IE 1573
QY 615 RPEKSDILNDYS0TSDIIGDKYGNSS0ITTKTAPPRSDNNDK0NSK0SLED0PANNE0LQ 674
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1574 ILEEC-IISAMPTKSSRAK0PA0TASKL-----PPVARKPS0L0PYKLLP0QNR-LQ 1625
QY 675 Q0LEVPHTKEDD-----SILANSSNIAPPEEL----- 701
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1626 P0KHVSFTPGDDMPRVYCV0EGTPINFSTAT0SLD0LTIESPPNELAAGEV0RGA0SGEPE 1685
QY 702 ---TLPV0EADNDYSFNDVTKTFDAYSSFEESISREHEDSKPINFISIMH0K0K0KH0 758
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1686 KRDTIPTG0RSTDE0G0K0T0SV-TIPELD0NK0AEG0ILACIN-----SAMPK0K0SK 1739
QY 759 IHKVP0K0I0IASY00-----YKNE---0ESRVTSDKVKI0PNAI0EKK0REVNVMSR- 806
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1740 PF0RV--KKIM0V0Q0AS0ASS0SAPNK0LDGKKK0PTSPVKPI0P0NTEV0T0R0K0NADS0 1797
QY 807 ---RVSP0MD0LNT0S0FL0PELSEDSGFK0LNFANYSNN0NR-----P0SFTPL 852
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1798 NLN0R0VSDNKD--SKR0NK0KNSKD--FND---KLPN0EDRV0GSAF0D0PHIYPI 1849
QY 853 ---STKVLNSI0ND0PVN0PEP0K0SYA0EIRNARLSANKA0PNA0APLP0Q0R0S 905
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1850 EGT0PCF0RND0SLSD0DD0VD0L0SREK--AELRKAKENK0E0K0V0SHTEL-----T 1901
QY 906 STRNSNKR0VSRFRVPT0EIR0TSSALAPCDMYNDI0FDEFGAS0K0PTI0KA0EMK0TLP0SMD 965
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1902 SNO0SANK0TAI0AK0P---INR-----G0QPKPII0K0--ST0P0SS 1937
QY 966 K0DVYRII0NAK0GY0DEYI0NAKLV0DKP-----KKN0SV--0D 1002
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1938 K0IPDR-----GAAT0EKL0N0FAL0ENT0P0CF0SHNS0LSS0LSDI0E0NNK0NEPIK0TE 1991

```

```

QY 1003 PEDRYEE-----LOOTASIHNAITDSIYGRPD0SDTMDLPTL 1040
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1992 PPD0SG0E0P0K0P0ASG0YAPK0SFHVED0TPVCF0SRSS0LSS0ID0E-----D0LIQ0ECI0SA 2046
QY 1041 S0ELK0PPTALL0ADRLEF0E0VH0PLR0NSVYVHPGAGAT0NSSML0PEP0FELINS0PARN 1100
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2047 MPK0KKRP-----SRLK0DNEKH0SPRNM0GIL--GEDTL0LDK0I0RPD0SEH0GLSPD-- 2095
QY 1101 V0NNS0NVAI0SGNASTI0S0LDMNFD0ATIG0KIQ0E0PASK0ANTV0GDD0GLASAP0 1160
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2096 -SEN0D0WKAI0EGANSI-----VSSLH0QAAA0CLSR0-ASSDS0SIL0SKG0ISL0GSP 2147
QY 1161 TP0PTPK0ESK0SPAKLS0SAPR0KPIK0SPVRYIKKNG0SIAGI0PIPKAT0HPK0K0F 1220
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2148 FLUT0P0E0---KRP-T0NK0GR-----ILK0PEK0STLE--TK0IES0K0I 2188
QY 1221 0GNEI0SNH0KVRD0GIS0PSSG0SEH0Q0HP-----SNV0SP0YTDA0T0IVPD0EN 1268
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2189 K0GK0K0YK0SLI0GK0VNS0EIS0G0MK0P0LQ0ANMP0IS0RGRTM0H0I0GV0RNS0SS0T0SPV0SK 2248
QY 1269 K 1269
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2249 K 2249

```

```

RESULT 13
US-10-022-819-2
; Sequence 2, Application US/10022819
; Publication No. US20030027166A1
; GENERAL INFORMATION:
; APPLICANT: ALLEN, Antonette C. P.
; OLSEN, Sheri J.
; LAWRENCE, Tammy
; ANGELLY, Tracy S.
; RABIN, Mark B.
; TITLE OF INVENTION: CODING SEQUENCE HAPLOTYPE OF THE HUMAN
; BRCAL GENE
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan Lewis & Bockius LLP
; STREET: 1111 Pennsylvania Avenue
; CITY: Washington DC
; STATE: District of Columbia
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible.
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/022,819
; FILING DATE: 22-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/074,452
; FILING DATE: 1998-05-06
; ATTORNEY/AGENT INFORMATION:
; NAME: <Unknown>
; REGISTRATION NUMBER: <Unknown>
; REFERENCE/DOCKET NUMBER: 044921-5049-01-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-739-3000
; TELEFAX: 202-739-3001
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO

```

```

: FRAGMENT TYPE: N-terminal
: ORIGINAL SOURCE:
: ORGANISM: HOMO SAPIENS
: STRAIN: BRCA1
: HAPLOTYPE: OM14
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT: 17
: MAP POSITION: 17q21
: SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-022-819-2

```

```

Query Match      3.18; Score 268.5; DB 9; Length 1863;
Best Local Similarity 18.66; Pred. No. 3e-06;
Matches 333; Conservative 243; Mismatches 604; Indels 607; Gaps 81;

```

```

18 QLOPOSSASAFNSPTKPLNFPRTNSKPSLDPNSSD-----TYSNEDDOEGKREKEDT 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
170 RIQPOKTSVYI-----ELGSDSEDTVKNATYCSVGDDQLQITTPGT 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
73 AFQTSFDR-----NFDLNSIDIOOTIQHO-----QOPOQOQOOLQOTDN 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
213 RDEISLSDAKKAACEF---SETDVTNTEHQPSNNDLNTTEKRAAERHPEKYQG--SSVSN 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
113 NLIDFSGTQMT-----STLDLTKONPTVDKVNENNAPTYINTSPKSKTKKATF- 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
269 LHVPCGNTHTASSLOHENSLLLTCKDRMNVKAEFCNKSQOPGLARSOHRMAGSKETC 328
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
163 -PKAPKVAFTVNTPEINHNPDNVVEEDQSOOK-----EDSVEPLIQ----- 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
329 NDRKTPS-----TEKKYDLNADPLCEKEMKOKLPCSEMPROTEDEVPTWLLNSIOKV 382
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
207 HQMKDPSQFNYSDE-----DTNASVPTPTPLHTTKPTFAQLLNKNEVNSEPE 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
383 NEW-----FSRSDELGSDSDHGESESNAKV-----ADVLQVLENY----- 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
255 ALTKMKLRENFNLSDDEKYNLSTPTNNNNKSNVSDMSHLQNLDAASKNTENTNHN 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 -----DEYSGSS--EKIDLLAS-----DPR-----BALICK--SERVHS 449
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
315 LSPALKAPKNDIENPLNLTNADISLRSSGSSQSLQSLRDNRYLESVPSPKKVNPGL 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
450 -----KVESNIEDKI-----FGKTYRKASLPLSHVTEULI----- 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
375 SLNDGKIGFSOE---VVESSLPRDLSRDKLETTEKENDAPRENNENFIDAKSTNTNGQL 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
483 -----IGAFVTEPOIIOERPLTNKLKRKRPTSLHP-----EDFI-----KKADLA 524
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
432 VSSDHLDSFDRSYNHTPOS---IINLNSASOSQSLALEKOR-----QTOEOQOTQAA 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
525 VOKTPEM--INOGTNOIEONGVOMNITNSGHENKTKGDSIONEKPNPDESLEKESAFKT 582
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
485 EPEETSFSDNIKV-----QEPKSN-----LEFVVTIKKRPVSATKKA 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
583 KAEPISISSIMLELNTNHNKAPKKNLRKRSRTHALELV--VSNLSPPNTELOI 641
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
526 PKRPFSSILRIKED-----ETAPADIIHK--KENANSHVEDTALLKKA 571
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
642 DSCSSSEIKKKKNOMPVHRSRMLQMEGKEPATGAKKSNKPNQOTSKRSDSTFPELK 701
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
572 LNDEESTPTONSTRKMSIRF-----HIDSDMKLED--SNDGREDNDOI-----SREK 618
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
702 LTNAPGSTKCSNTSELKEFVNPSPLEPEKEKLETVKVSNNNA--EDPKDLMLGSEVLQ 759
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
619 SDILNDVQSDIIGDKYG--NSSSEITTKTLAPPSSDNDN-----KENSSTL----- 664
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
760 TERVESSISLVPGTODGTGTOESISLEVTLGAKATEPNKCVSCGAFENPKGLIHCS 819
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
665 EDPANN--ESLQOOL--EVNPTKEDDSILANSNIAPEPELTLPVVEANDVSSFN----- 715
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
820 KDNNDTGEFKYPLGHEVNHRSIEMEESELDQYLQNTFFKYSKRSQSFALFSPGNAE 879
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
716 DVTTFDAVSSFEESLSREHETDSKPINFISIMHKOEKQKHQ--IHKVPTKQIIIASIQQ 773
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 880 ECATFSASHSG---SLKQO---SEKVTF--EEOKEENOGKSNESNIKPEVQTVNITAGF-P 930
Qy 774 YKNEQESRVTSQKVIKIPAIQOFKKEEVNVMRSRVSPDMODLVNSQF--LPELSDSEF- 831
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 931 VVGQKDKFVDNAKCKSTKGSRCFLSSQFRGNETGLITNNKIGLONPRIRPLPPIKSFV 990
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 832 -----KDL---NFANYS-----NNTNRRPSFTPLSTKNVLSNIDNPVNEPPEPKSY 876
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 991 KTKCKKNLLENFEEHSMSPEREMENENIPSTVSTISNNIRNV----- 1035
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 877 AEIRNARLSANKAAPNAPPLPQROPSTRSNSNKNVSRFRVTFEIRTSALACD 936
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1036 -----FKGASSNINEVSGSTNVGSSINIGSSD 1065
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 937 MYNDIFDEFGAGSKPTIKAEQKTLPSMDKDDVKKLNAKKGVQDDEVYINKLYDQKKK 996
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1066 --ENIQAEELGRNREKLLNA-----MLRGLVLOPEVYKQSLPESNCKH 1105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 997 NSIVTDPEDRYEELQOTASIHNATIDSSIGRPDSISTDMPLYL--SDELRKPTALLSAD 1055
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1106 PEI---KQOEYEEVYQT-----VNTDFSPYILSDMLEP----- 1136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1056 RLFMEQEVNPLRNSVLVHPGAGATNSMLPEPDFELINSPARVNSNDSNVAISGNAS 1115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1137 -----MGSSHASQVQSETPDDL-----DQGEIKEDT- 1163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1116 TTSFNOLDMPNDQATIGOKIOEOPASKS-----ANTVRGDDDLASAP 1159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1164 --SFAENIK--ESSAVFESKSVQRGELSPSPFTTHLAOGYRRKAKLESSEULSS-- 1218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1160 ETPRTPTKRESISSKPAKLSASAPRSPKIKGISPVYIKNGSIAGIPIKATHKPKKS 1219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1219 EDEELPCQHLFLGKVNNIPQOSTRSTV-----ATCLSKTFEENLLS 1262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1220 FQG--NEISNHNKVRDGGISPSGSEHQOHNPSMVP--SQYTDATSTVPDENKQVQHKPR 1276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1263 LKNSLNDCSNOVILLAKASQEHNLSEETKCSASLPSQSCSELEDLTANTNTDPELIGSK 1322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1277 EKOQKNNHNNHNNHNNQO---KTDIRGVVDEIDPVGLOERGKLEFRVLGIKINILPDIN 1333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1323 QMR-----HQSESGVGLSDKELYSDDEBERTGILENNQO----- 1356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1334 THKGRFTLLDNGVCHVTTPRENMDDH--NVAIGREFELTYAASLEFILLTKASYEKPRG 1391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1357 -----EEOQSMDSNLGEAASGCESETSVSED----- 1381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1392 TLVEYTEKKVVKSRNRSRLGSKDILITTKFVPTREVDTMANKFAPGSPARCVIDLOQ 1451
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1382 -----CGSLSSQSDILT-----TQORDTQOHNL-----IKLOQ 1409
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1452 FEDQITGKASQFDLNCFFNEW-----ETMSNGNO-----PMKRGKPYKIAOLE 1493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1410 EMALAEAVLEQHGSPKSYSTIISDSSALDRLNRPDSTSEKAVLTISQKSEVPISSO- 1467
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1494 VKMLVPRSDPRELIPISIRAYESINELNNDQNNYFEGYILHQEGGDCPTTKKPEFKMG 1553
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1468 -----NPEGLSADFEVASDSTSKNKE---PGVRSRSPSKCPSLDDRWYMHSC 1513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1554 TSLAHSEISIKTRAKINLSKVVDLIYVDKENIDNSNHRNSDYL 1600
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1514 SGSLOANRP-----SQEELIKVVD--VEEOULEBSGPHDLTETSYL 1553
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 14

US-09-734-672-4
Sequence 4, Application US/09734672
Publication No. US20020183268A1

GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.
Allan, Antonette C.
Alvares, Christopher P.
Critz, Brenda S.
Olson, Sheri J.

Schelter, Denise B.
Zeng, Bin
TITLE OF INVENTION: Coding Sequences of the Human
BRCA1 Gene
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan Lewis & Bockius LLP
STREET: 1111 Pennsylvania Ave., N.W.
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/734,672
FILING DATE: 03-Dec-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/966,436
FILING DATE: 07-Nov-97
APPLICATION NUMBER: US 2002/0183268A1-97
FILING DATE: 12-Feb-96
ATTORNEY/AGENT INFORMATION:
NAME: Michael S. Tuscan
REGISTRATION NUMBER: 43,210
REFERENCE/DOCKET NUMBER: 44921-5055-02-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-739-3000
TELEFAX: 202-739-3000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: No. US20020183268A1 Relevant
TOPOLOGY: No. US20020183268A1 Relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCA1
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-734-672-4
Query Match 3.1%; Score 267.5; DB 9; Length 1863;
Best Local Similarity 18.8%; Pred. No. 3.4e-06;
Matches 336; Conservative 243; Mismatches 601; Indels 607; Gaps 82;
QY 18 QLOPSSASISFPTKPLNFPRTNSKPSLDPNSSD-----YTSDDDEKGEKKKDT 72
DB 170 RIPOKTSYI-----ELGSDSSEDTYVKATYCSVGDELLQITPOGT 212
QY 73 APTSTDR-----NPDLSIDIQOTIOH-----OOQPOQOOLQOTDN 112
DB 213 RDEISLDSAKKACF---SETDYTNTEHHQPSNNDLNTTEKRAAEHPEKYQG--SSVSN 268
QY 113 NLIDESFOTPMT-----STLDLTQONPTVDKY---NENHAPTYINTSPKSIIMKAT- 162
DB 269 LHYEPGCTNTHASSLQIHNSSLITFKDRMVEKAKEPCNKSKOPCLARSQHNRMAGSKETC 328
QY 163 -PKASPKKAFYVYNPEIHHYPDNRVEEDOSQOK-----EDSVPEPLIQ----- 206
DB 329 NDRRTES-----TEKVDLNADPLCEKKEWKKQKLCSENPRTDEVPWITLSSIOKV 382
QY 207 HOKKDSQEFYSD-----DTNASVPTPTPLHTTKPTFAOLKNNNNVENSEPE 254
DB 383 NEW-----FSRSDLLGSDSDSHGSESNKV-----ADVLDLNEV----- 419
QY 255 ALIDMKLRNFNSLIDKVNLYLSPYNNNSKNVSDMDSHLQNLQDASKNKTNENIHN 314

DB 420 -----DEYSGSS--EKIDLAS-----DPH-----EALICK--SERHVS 449
QY 315 LSFALKAPKNDIENPLNSLINAIDSLSSGSSQSSLOSRLRDNRYLVESPOSPKKVNBGL 374
DB 450 -----KSVESNIEDKI-----FGTYRKKASLPMLSHVTENLI----- 482
QY 375 SLNDGIGFQDE---VVESSLPRDSRDKLETTTEHAPAEHNNNPFIDANKSTNNKQGL 431
DB 483 -----ICAFYTEPOIIOERPLTNKLRKRRTSGLHP-----EDFI-----KKADLA 524
QY 432 VSSDHLDSFDRSYNHTEOS--ILNLSNASQSOISINALEKOR-----QTOEOEOQAA 484
DB 525 YQKTPEM--INGTNGTEQNGQVANNITNSGHENTKTGSGIQENKNPNPISLEKESAFT 582
QY 485 EPEETSFSDNIKKY-----QEPKS-----LEFYKYITKKPVSATEIKA 525
DB 583 KAEPISSISNMELELNHNSKAPKKNLRKRKSTRHIALELY-VSRNLSPPMCTELQI 641
QY 526 PKRESSRIILRKMD-----ELAEPADHPK--KENDANSHEVDTALLKKA 571
DB 642 DSCSSSEIIRKKKYNQMPVRHSRNLOLMEGKEPATGAKSKKPMEDOTSKRHDSOTFPELK 701
QY 572 LNDEESDTQNSTKMSIRF-----HIDSDMKLED--SNDGDREDNDI-----SREK 618
DB 702 LTNAPGFTKCSNTSELKEFYNPSPLPREKKEKLETYKVSNN--EDPKDMLGGERLYQ 759
QY 619 SDILNDYSQTSIDITGDKYG--NSSSEITTKTLAPPRSDND-----KENSKL----- 664
DB 760 TERSESSSISLVPGTGYDOESISLLEVSTLGAKRKEPNKVCSCAFAFEPKGLIHGCS 819
QY 665 EDPANN--ESLQOOL--EVPHTKEDDSILANSNIAPEELLPVEANDYSFN----- 715
DB 820 KDNNDNDEGEFYPLGHEYNHRETSIEMESELDAQYIQTFFKSKRQSFALFSQNAE 879
QY 716 DVTKTFDAYSSFEESLSREHETDSKPINFISIMHKOEKOKHQ--IHKVPKQI1IASYQ 773
DB 880 EECATFSAHSG--SLKKQ--SPKVTF--CEQKEENQKNESNIKPVQVINITAGR--P 930
QY 774 YKNOESRYTSDKYKIRPAIOFKFKFEYNVNSRRVSPDMODLVNSQF--LELSDSQF- 831
DB 931 VVGQKDRPYVDNAKCSINGSRFLCSSQFRGNETGLIPNKGILLQNPRIPLPPIKSFV 990
QY 832 -----KDL--NFANYS-----NNTNRPSTPLSTKYNLSINDNDPNVVEPEPKSY 876
DB 991 KTKCKKNLLENFEBHSSPREKGENENIPSTVSTISNNIRENVFK----- 1038
QY 877 AEIRNARLSANKAPNQAPLPPOROPSSSTRSNSNKRVSFRVPTFEIRRTSALACD 936
DB 1039 -----ASSNINEVGS-----STNEVGSSINELIGSSD 1065
QY 937 MYNDIFDDFGAGSKPTIKABEMKTLPSMDKDDVKRILNAKGVYODEIYNAKLYDQKPK 996
DB 1066 --ENIQAEELGRNRGPKLNA-----MLRGLVLOPEVYKOSLPESNCKH 1105
QY 997 NSIYTDPEDRYEELQOQASIHATIDSISYGRPOSISTDMLPYL--SDELKRPYALLSAD 1055
DB 1106 PET---KQOYEYVYOT-----VNDPFSPLYLSDMLEOP----- 1136
QY 1056 RLFMEOEYHPLRNSVYLHPGAGATNSSMLPEPDFELINSPARVNSNSDVAISGNAS 1115
DB 1137 -----MGSSHAQVCEPTDDLL-----DDGEIKEDT- 1163
QY 1116 TISFNOLDAMPDDOATIQKIOEOPAKS-----ANTVRGDDGLASAP 1159
DB 1164 --SEAENDIK--ESSAVFSKSYQKGLSPSPFTHTHLAGYRQAKKLESSEENLIS- 1218
QY 1160 ETPTPTTKESISSPAKLSASPKSPKIKIGSPRVYIKKNGSITAGIPKATPKPKS 1219
DB 1219 EDEELPCQHLHFGKVNINIPQOSTRHSTV-----ATCLSKNTEENILS 1262
QY 1220 FQG--NEISNHRKVDGIGTSPSSGSEHQOHNSMNSVP--SQYTDATSTVPDENKDVQHKPR 1276

```

Db 1263 LKNSLDCSNQVILAKASQEHHLSEETKCSASLFSQSCSELEDLTANTNTODPFLIGSSK 1322
Oy 1272 EKOKKHNRHHHHHHHKO---KTDPGVYDDEIDPVGLOERGLFPRVLGIKINILPDIN 1333
Db 1323 QMR-----HOSRQGVGLSDKELVADDEBERGIGLENNQ----- 1356
Oy 1334 THKGRFTLLDNGVHCVTTPPEYNMDH--NVAIGKEFELTVADSLFEILLKASYEKPRG 1391
Db 1357 -----EEOQMSDNLGEAASGCESTSVSD----- 1381
Oy 1392 TLVEVTEKVVASRNRLSLFGSKDIITTKFVPTVKOTWANKFAPDGSFARCYIDLQO 1451
Db 1382 -----CSGLSSQSDILT-----TOQRDTMOHNL----- 1409
Oy 1452 FEDQITGKASQDFLNCFNW-----ETMKNQNO-----PMKRGPKYKIAOLE 1493
Db 1410 EKALELVLEQHGQSPNSNPTSISSSALDLRNEQSTSEKAVLTSQKSYPTISQ-- 1467
Oy 1494 VKMLVYPRSDPREILFTSIRSAVESINELNNEONNYFEGYLQEGGDCPTFKKRFKRLMG 1553
Db 1468 -----NPEGLSADKFEVASDSTSKNKE-----PQVERSSPSKCPSLDRWYMHSC 1513
Oy 1554 TSLAHSEISHKTRAKINSKVVDLYVDKENIDRSNHNHNSDVL 1600
Db 1514 SCSLQNRNP-----SOBELIKVVD--VEEQOLESGPHDLTETSYL 1553

```

```

RESULT 15
US-09-982-828-6
; Sequence 6, Application US/09982828
; Publication No. US20030022184A1
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
;           Allen, Antonette C.
;           Alvarez, Christopher P.
;           Critz, Brenda S.
;           Olson, Sheri J.
;           Thuerber, Denise
;           Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
;                   BRCAL Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESS: Morgan Lewis & Bockius LLP
; STREET: 1111 Pennsylvania Avenue N. W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/982,828
; FILING DATE: 22-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/074,453
; FILING DATE: 1998-05-06
; APPLICATION NUMBER: US 08/738,691
; FILING DATE: 1997-02-12
; APPLICATION NUMBER: US 08/598,591
; FILING DATE: 1996-02-12
; ATTORNEY/AGENT INFORMATION:
; NAME: Michael S. Tuscan
; REGISTRATION NUMBER: 43,210
; REFERENCE/DOCKET NUMBER: 44921-5053-01-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-739-3000
; TELEFAX: 202-739-3001
; INFORMATION FOR SEQ ID NO: 6:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCAL (om13)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-982-828-6

Query Match 3 1%; Score 267.5; DB 9; Length 1863;
Best Local Similarity 18.8%; Pred. No. 3.4e-06;
Matches 336; Conservative 243; Mismatches 601; Indels 607; Gaps 82;

Oy 18 QLDQSSASIFNSPYKPLNFPRTNSKPSLDPNSSSD-----TYTSEQDQEKKEKEDPT 72
Db 170 RIQPKQTSYVI-----ELGSDSSSDPTVKNKATYCSVGQQLLQITPGQT 212
Oy 73 ARQSTFDR-----NFDLNSIDIQOTIOH-----QQPQQQQQLSQTGN 112
Db 213 RDEISLDSAKKACEF-----SETDVTNTEHOPSNDLNTTEKRAERHPEKYOG--SSVAN 268
Oy 113 NLIDFSPQTPMT-----STLDLTQNPYQKV--NENHAPTYINTSPKNSIMKAT- 162
Db 269 LHYEPCGTTTHASSLOHENSLLITKDRMNVKAEFCNMSKOPGLARSOHNRAGSKETC 328
Oy 163 -PKASPKVAFTVTPNPEIHHPDNRVEEDQSOQ-----EDSVEPPLIQ----- 206
Db 329 NDRRTGS-----TEKKVLDLNDPLCEKKEWKKPLPCEENPRDPTDVMITLNSIQKY 382
Oy 207 HQMKDPSQFNYSDE-----DTNASVPTPTPLTTKPTFAQLLNKNEVSEPE 254
Db 383 NEW-----FSRSDDELGSDSDSHDGESESNKV-----ADVLDVINEV----- 419
Oy 255 ALIDMKLRKFNFSNLSDKENVLYLPTNNNSKNVSDMSHLQNLQDSKKNKTNENIN 314
Db 420 -----DEYSGSS--EKIDLLAS-----DPI-----EALICK-SRVHS 449
Oy 315 LSFALKAPKNDIENPLNSLTNADISLRSSGSSQSSLSQSLRNDNRVLESVPGSKVNPGL 374
Db 450 -----KSVESNIEDKI-----PGKTYRKKAASLPNLSHYVENLI----- 482
Oy 375 SLNDGIKGFSD-----VESLPLPDLSRDKLETTKEKADAEHNNEINTDAKSTYTKGOLL 431
Db 483 -----IGAFTVEPOLIQRPLTNKLKRRRPTSGLHP-----EDFI-----KKADLA 524
Oy 432 VSSDDHLDFDRSYNHTEOS--ILNLTNSASQSOISLNLKOR-----QTOBOEQTOAA 484
Db 525 VQKTPEM--INQGTNOTEDNGOVAMITNNGHEKTKGDSIQNEKNPNPLESLEKEAFET 582
Oy 485 EPEETSFSDNIKVK-----QEPKSN-----LEFVKYTIKKEPVASATEIKA 525
Db 583 KAPISSISNMLELNLHNSKAPKRRKSTRHIALELY--VSRLSPENCETLOI 641
Oy 526 PKREFSSRLIRIKNED-----ELAFADLHPK--KENENASHVEDDLDALLKKA 571
Db 642 DSCSSSEETIKKKRYNOMPVHRSHNLQLMGKKEBATGAKKSNNPNEQTSKRHSDPTPELK 701
Oy 572 LNDEESDTONSTKMSIRF-----HIDSMDKLED-----SNDGRDRNDNDI-----SRFK 618
Db 702 LITAPGSFTKCSNTSLKEFVNPSLPREKEKELLEVYKXSNN--EDPADMLMSGRVQLQ 759
Oy 619 SDILANDVSQTSILIGDKYG--NSSSEITTKTLAPPSSDND-----KENSLSL----- 664
Db 760 TEKSVSSISIVPGTDYQGESISLLEVSTLGKAKTEPNKVCQSCAARENPGILHGCS 819
Oy 665 EDPANN-ESLQOOL--EVPTTKEDDSILANSSAIAPPEELITLPVYANDYSSFN----- 715

```


Db 820 KDNRRNDTEGFKYPLGHEVNHRSRETSIEMESELDQAQYLQNTFKVSKROSFALFSPGNAE 879
QY 716 DYTTFDAYSFPEESLSREHETDSKPIINFISIMHKOEKKHO--IHKVPTKOIIASYQ 773
Db 880 EECATFSAHSG---SLKKQ---SPKYTF--ECEQKEENQCKNESNIKPVQIVNITAGF-P 930
QY 774 YKNEOSRYTSDVKIIPNAIOFKFKFEVNVMSRRVSPMDDLNVSOQ-LPELSEDSGF- 831
Db 931 VVGOKDKPYDNMAKCSIKGSRFCLSSQFRGNENGLITPNKHGLQNPYRIPPLFPFKSFV 990
QY 832 ----KDL---NFANTS-----NNTNRRSFPTPLSTKNVLSNIDNDPVNVEPPEPKSY 876
Db 991 KTKCKKNLLEENFEHESMSPEREGNMENIPSTVISRNNIRENVFKE----- 1038
QY 877 AEIRNARLSANKAPRQAPPLPQROPSSSTRSNMKNRVRFRVPTFEIRTSALAPCD 936
Db 1039 -----ASSNINEVG---STNEVGSSINEIGSSD 1065
QY 937 MYNDIFDDFGAGSKPTIKAGMKTLPSMDKDYKRIINAKKGYTODEYINAKLVQCKPKK 996
Db 1066 --ENIQAEIGRNNGPKLNA-----MLRLGVLPQPEVYKOSLPGSCKH 1105
QY 997 NSIVTPDEDRYEELQOTASTIHNAITDSIYGRPDSISTDMLPYL-SDELKKPTALLSAD 1055
Db 1106 PEI---KQXEYEVVOT-----VNTDFSPYLLISDNLEQ----- 1136
QY 1056 RLMEQEVHPLRSNSVLVHPGACATNSMLPEPPFELINSPARVNSNDSNVAISGNAS 1115
Db 1137 -----MGSNASQVCESTPDDL-----DDGEIKEDT- 1163
QY 1116 TISFNOLDNMFDDOATIGOKIOEOPASKS-----ANTVRGDDGLASAP 1159
Db 1164 --SFAENDIK--ESSAVFSKSVQKGELSRSPFTHTHLAGYRRGAKKLESSEENLSS-- 1218
QY 1160 ETPRTPTKKSISSKPAKLSASAPKSPKIKGSEVRIKKNGSTAGTEPIPKATHPKPKS 1219
Db 1219 EDEELPCFOHLLFGKVNINIPSOSTRHSTV-----ATECLSKNTEENILS 1262
QY 1220 FOG--NEISNHNKVRDGISSPSSGEHQHNPMSVVP--SOYDATSTVPDENKDVQHKPR 1276
Db 1263 LKNSLNDCSNOVILAKASQEHNLSEETKCSASLFSQCSLEEDLTANTNTQDPFLIGSSK 1322
QY 1277 EKOKOKHHHHHHHHKQ---KTDIPGVYDEIIPDVGLQERGLFFRVLGIKNINILPDIN 1333
Db 1323 QMR-----HQSSEQVGLSDKELVSDDEERGTLLENQ----- 1356
QY 1334 THKGRFTLLDNGVHCYTPTEYIMDDH--NVAIGKEPELTVADSLFILTAKASYEKPGR 1391
Db 1357 -----EOGMDSNLGEAASGCESETSVSED----- 1381
QY 1392 TLVEYTEKKVVKSRNRLSRLFGSKDILITTKFYPTVEKDWANKFAPDGSFARCYIDLQ 1451
Db 1382 -----CSGLSSQSDILT-----TOQRDTMQHNL-----IKLQ 1409
QY 1452 FEDOITGKASQFDLNCFNEM-----ETMSNGNQ-----PMKRGRPKYIAQLE 1493
Db 1410 EMAELEAVLEHQSGSPNSVPSIISDSALEDLRNPQOSTSEKAVLTSQKSSEYPIQ-- 1467
QY 1494 VKMLYVPRSDPREILPTLSISAVESINELNNEQNNYFEGYLDHOGGDCPIFKKREKLMG 1553
Db 1468 -----NPEGLSADKFEVSADSTSKNKE-----PGEVRSPPSKCPSLDDRWYMHSC 1513
QY 1554 TSLLAHSEISHKTRAKINLSKVVDLIYVDKENIDRSNHRNPSOVL 1600
Db 1514 SGSLQNRNRP---SOBELIKYVD---VEQQLSESGPHDLETSTYL 1553

Search completed: March 17, 2003, 12:34:17
Job time : 93.588 secs

This Page Blank (uspto)

Shears, Beverly

From: Devi, Sarvamangala
Sent: Tuesday, March 25, 2003 7:02 AM
To: Shears, Beverly
Subject: RE: 09/964,858

Good morning Beverly:

Sorry Beverly. I have not received sequence search report for this case. I would appreciate if could please perform a sequence and an interference search for SEQ ID NO: 1 or print out the sequence report from the March 17th search, if it was saved. Thanks a lot.

-----Original Message-----

89182

Sh ars, Beverly

From: Devi, Sarvamangala
Sent: Friday, March 14, 2003 4:38 PM
T : Shears, Beverly
Subject: 09/964,858

Hello Beverly:

Would you please perform a sequence and an interference search for SEQ ID NO: 1 and a fragment comprising amino acid residues 1-263 of SEQ ID NO: 1 in case 09/964,858?

Thanks.

S. DEVI, Ph.D.
AU 1645
CM1-7E15

7E12
mb

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 1E05 Tel: 308-4994

This Page Blank (uspto)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:19:45 ; Search time 9.28075 Seconds
(without alignments)
5839.012 Million cell updates/sec

Title: US-09-964-858-1_COPY_1_263
Perfect score: 1386
Sequence: 1 MNSPKLPIIDKSHLQ.....NKNNEVSEPALIDMKLR 263

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

arched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	158.5	11.4	1709	5	095PH5
2	146	10.5	542	5	096600
3	145.5	10.5	2454	5	087263
4	142.5	10.4	1721	5	088500
5	142.5	10.3	2678	5	09NDS4
6	142	10.2	2843	5	096315
7	138.5	10.1	734	5	0871A0
8	138.5	10.0	1166	2	086489
9	137	9.9	730	5	087867
10	137	9.9	756	11	09QY60
11	137	9.9	947	5	015739
12	135.5	9.8	2472	5	0872M5
13	134.5	9.7	1162	4	096JX9
14	134.5	9.7	1171	2	09KWX6
15	134	9.7	1141	16	099W46
16	134	9.7	1141	16	0932F7

17	133.5	9.6	800	5	096902	096902 dictyosteli
18	133	9.6	628	5	0960L0	0960L0 drosophila
19	133	9.6	767	5	09V9Z6	09V9Z6 drosophila
20	131.5	9.5	1179	12	091L98	091L98 white spot
21	131.5	9.5	1180	12	08VAS9	08VAS9 white spot
22	131.5	9.5	2439	5	09VMS2	09VMS2 drosophila
23	130.5	9.4	1140	5	0871T0	0871T0 dictyosteli
24	129.5	9.3	803	5	09BLH2	09BLH2 halocynthia
25	128	9.2	548	5	0871H3	0871H3 dictyosteli
26	127.5	9.2	1335	2	033635	033635 staphylococ
27	127.5	9.2	1542	5	09VE07	09VE07 drosophila
28	127.5	9.2	1789	5	0872E3	0872E3 dictyosteli
29	127	9.2	736	3	074500	074500 schizosacch
30	127	9.2	816	5	0872N6	0872N6 dictyosteli
31	127	9.2	1338	5	077306	077306 plasmodium
32	127	9.2	1483	5	087273	087273 dictyosteli
33	126.5	9.1	601	3	09C1I3	09C1I3 candida alb
34	126.5	9.1	660	5	094485	094485 dictyosteli
35	126	9.1	1593	5	0871T7	0871T7 dictyosteli
36	125.5	9.1	838	5	09Y0C9	09Y0C9 dictyosteli
37	125	9.0	641	5	061640	061640 drosophila
38	125	9.0	646	16	09KMZ5	09KMZ5 vibrio chol
39	125	9.0	720	5	023847	023847 dictyosteli
40	125	9.0	722	5	09VEK4	09VEK4 drosophila
41	125	9.0	1545	5	09YVNI	09YVNI drosophila
42	124.5	9.0	556	5	095RC6	095RC6 drosophila
43	124.5	9.0	773	16	09SR04	09SR04 mycoplasma
44	124.5	9.0	811	5	09VT40	09VT40 drosophila
45	124.5	9.0	870	13	09PTB3	09PTB3 colurnix co

ALIGNMENTS

RESULT 1

095PH5 PRELIMINARY; PRT; 1709 AA.

AC 095PH5: 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DT Histidine kinase DhKL.

GN DHKL.

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.

OX NCBI_TaxID=44689;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AX4;

RA Anjard C., Loomis W.F.;

RT "The histidine kinases of Dictyostelium.";

RL (In) Inouye M., Dutta R. (eds.);

RL Histidine kinases in Signal Transduction, pp.1-1, Academic Press,

RN San Diego (2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=AX4;

RA Anjard C., Loomis W.F.;

RT Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF362373; AAK54092.2; -

KW kinase.

SO SEQUENCE 1709 AA; 192603 MW; F0A91C505D5DE178 CRC64;

Query Match 11.4%; Score 158.5; DB 5; Length 1709;

Best Local Similarity 26.4%; Pred. No. 0.0078;

Matches 71; Conservative 34; Mismatches 127; Indels 37; Gaps 8;

QY 2 MNSPKLPIIDKSHLQDP-----QSSASIFNSPTKRLNPPRTNKRSLDPNSSDPT 55

DB 456 MNSPSSHTP---NSPMIFQPIVSNICNSGSGSNNSPHHIDN-----NNSNQOO 502

QY 56 YTSEQDDEKGEKKEDTAFQTSFDRNFDLNSIDI---QDTIOHQDQDPPQDQDOLQDTON 112

[illegible]

RESULT 2	
0966U0	
ID 0966U0	PRELIMINARY;
AC 0966U0	PRT; 542 AA

DT	01-DEC-2001 (TREMBlrel. 19, Created)
DT	01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DATA	

Dictyostellium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium
 NCBI_TaxID=44689;

RP SEQUENCE FROM N.A.
RX MEDLINE=21256138; Pubmed=11356336;
RA Epub 11 Nov 2001

MA Tsuruji A., Kodaira K., Inoue M., Yasukawa H.;
RT "Endonuclease IV homolog from *Dictyostelium discoideum*: sequencing and
functional expression in AP endonuclease-deficient *Escherichia coli*."; *Mol.
Mutat. Res.* 486:53-57(2001).

```
DR EMBL; AB055424; BAB59036.1; -.
DR InterPro; IPR001719; AP_endonuclease2
DR Pfam; PF01261; AP_endonuc_2; 1.
```

DR	PROSITE: PS00729; AP_NUCLEASE_F2_1; UNKNOWN_1.
DR	PROSITE: PS00730; AP_NUCLEASE_F2_2; UNKNOWN_1.
DR	PROSITE: PS00731; AP_NUCLEASE_F2_3; UNKNOWN_1.

SEQUENCE 542 AA; 61952 MW; 28CFAFBA699CEBE3 CRC64;

Query Match	10.5%	Score 146;	DB 5;	Length 542;
Best Local Similarity	24.6%	Pred. No. 0.017;		
Matches 69;	Conservative 41;	Mismatches 104;	Indels 66;	Gaps 12

```

Qy 22 QSSASLFSPTKP---LNPRTNSKPSLDPNSSSDYTSSEQDEKKG-----EKK 70
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 QHSKQQLFLNNKKNILNF---NSK-----LYSTTYKNNNNNNNNNOILNKTIKRKN 61

```

```

1 DFAQSFSD-----RNFOLDNSIDIQTI-----QHQQQQPQQQQQQLSQGTBN 112
: 11 1 : : : : 1 1 1 1 1 1 1
62 NEITQFEEDDDNDNGSIHDCDSCDEVDNEEIIIPESPSPSKLQQQQQQQQQSKLPKQT 121

```

```

Qy 113 NLIDFESFQTPMISTLDTKONPTVDKVNENHAPPIYINTSPKSKIMKATPKASPKVAF 17Z
    :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122 SITDFS---PIVKTISKADKISPEFS--NVNNASITNTTNNKNVKNKKTITTTTTRK--- 173

```

```

QY 173 TVTNPELTHYPCNRVEEEDDSQOKEDSYERPLIQHWKDPSEQFNYSDEDTNA----SVPP 228
      | : : |||: ::|| : : : : |
Db 174 --RNNKDEENEDDNEEEEEEEFDK-----KSKKKKTTTTTTTTTAYKKKKSSDP 232

```

```
OY      229 TTPBLHTTKPTFA-----QLLNKNNEVNSEPEALTDMLK 261  
           :: | : | : | : | :  
Db     224 KKKVNPTEKFSTINKDDQYVQKEEYSSSKKEKLNVKI 263
```

RESULT 3
Q8T2G3
ID 08T2G3 PRELIMINARY; PRT; 2454 AA

DT	
01-JUN-2002	(Tremblrel. 21, Created)
01-JUN-2002	(Tremblrel. 21, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
DE Hypothetical 274.1 kDa protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP
RR
RS SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guiso R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Neigel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL Submitted (MAR-2002) to the EMBL/Genbank/DBD databases.
DR EMBL; AC115608; AAL92375.1; -
KW Hypothetical protein.
SQ SEQUENCE 2454 AA; 274123 MW; B594001B4487C724 CRC64;

Query Match	10.5%	Score 145.5	DB 5	Length 2454
Best Local Similarity	24.0%	Pred. No. 0.092		
Matches	75	Conservative	30	Mismatches 120; Indels 87; Gaps 11;

```
QY      1 MNSTPSKLLPIDKSHLQLQPSSASAFNSPTAKLNFPRTNPKSLPDNS-----SSDT 55
       :||| | : : || | :||| | : || | : ||| | : ||| | : ||| |
Db     196 INST-----PVVQSQOOL--SNSSSLPTGTSPSLQSSSSSSSSSAAATASTATTSTT 248
```

```
OY      56 YTS0DQEKKEKKDTAFQTSTFDRNFDLDSIDIQOTIQHQQQQPQQQQQLSQTNNLI 115
```

QY	116	DEFS-----	FQTPMTSLDLTKQ-----	134
----	-----	-----------	--------------------	-----

Db 306 HEFQSIQIPCWDDGVP PPPPQGTFFENLTSVPIRPPPLFQHQQSTTSNNNNNTIIQQS 365

Qy 135 -----PTVDKVNENHAPTYINTSPNKSIMKATPKASPKKVAFTVTNP 177

```
Db      366 NTKSGDVTPTSLISPISTITNNNNNVNSTNTPAPTPLNPSSVTSPTS-----TSTNP 419
```

```

      : : : | : : : | : :
Db 420 PSNNPKPSIGQIOSLHYHNSLYQTPLFN--RNRGNNNNNNNDSTSSPMDSPLISS 476

```

QY	236	KPTFAQLLNKNN	247
Db	477	--TVAS-LNSNN	485

RESULT 4
Q8SSQ0
ID Q8SSQ0 PRELIMINARY; PRT; 1721 AA

DT	01-JUN-2002	(TREMBlrel. 21, Created)
DT	01-JUN-2002	(TREMBlrel. 21, Last sequence update)
DT	01-JUN-2002	(TREMBlrel. 21, Last annotation update)

OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;

AN
12]
RP
RC
STRAIN=AX4;
BA
Cloeckner G. Eichinger I. Czafarski K. Dachebat

RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.",
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

SQ SEQUENCE 1721 AA; 193133 MW; 68EA5A422CAFC8EC CRC64;

Query Match	10.4%	Score 143.5	DB 5	Length 1721
Best Local Similarity	25.2%	Pred. No. 0.086		
Matches 78; Conservative	30;	Mismatches 102;	Indels 99;	Gaps 15;

QY 18 QLOPOSSASIENS-----PTKPLNFR-----TNSKPSLDPNSSSDTYTSE 59

```

Db      524 QLOSTSSSSSSSSSSLSNLSNLSTQLPQATTTTTTTATSPSTSTSTSTSPSTSSLSLSQ 583
Qy      60 QDEKGEKKKDTAFQTSFDRNFDLNSLDIQOTIOHQOQOPOQOQO----- 106
Db      584 DKQLRREKHOIYIPK-----DKDSVANOOOQOQOQOQNGTTPSPNNESVT 631
Qy      107 -----LSOTDN-NLIDF-----SFQPMSTLIDLKQNTV-----DKN 141
Db      632 AATTTTSTNASTTTNNNTIPNFPNKKQLPATPTTGTPST--PTPQTSTSDNKON 689
Qy      142 E-NHAPYINTSP-----NKSIMKKATP-----KASPKVAFVTPTPEIHNPDR 186
Db      690 ENNNKMFVQKQTKLGKLFKSPVHKRTPPLGFGDSNSSSPSNNSMT--TSSSHS----- 743
Qy      187 VEEDSOQKQEDSVEPPLIOHQKDSQFNY-----SDEDTNASVPPPLHTTKPTFAOL 242
Db      744 -STNSSPMETSPGVESPKIT--KSPQONNILDVSDIGDNOOQOQOQDLTTTSTTTT 799
Qy      243 LKNNEVNS 251
Db      800 INNNNNNNN 808

```

RESULT 5		
ID	Q9NDS4	PRELIMINARY; PRT; 2678 AA.
AC	Q9NDS4;	
DT	01-OCT-2000 (TREMBLrel. 15, Created)	
DT	01-Oct-2000 (TREMBLrel. 15, last sequence update)	
DT	01-MAR-2002 (TREMBLrel. 20, last annotation update)	
DE	AmB.	
GN	AMIB.	
OS	Dictyostelium discoideum (Slime mold).	
OC	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.	
OX	NCBI_TaxID=44689;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=AX2;	
RC	MEDLINE=20118106; PubMed=10651904;	
RA	Kon T., Adachi H., Sutoh K.;	
RT	"AmB, a novel gene required for the growth/differentiation transition	
RL	in Dictyostelium.";	
RL	Genes Cells 5:43-55(2000).	
DR	EMBL; AB030033; BAB01489.1;	
DR	Interpro; IPR000194; ATPase_A/centre.	
DR	Prosite; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1	
NO	SEQUENCE 2678 AA; 298521 MW; 17D82BFBFB57C/CA CRC64;	

Query Match	10.3%;	Score 142.5;	DB 5;	Length 2678;
Best Local Similarity	20.9%;	Pred. No. 0.16;		
Matches 61; Conservative	48;	Mismatches 110;	Indels 73;	Gaps 10;

```

OY 2 NSTPSKLLPIDKHSHLQLOPOSSASAFJNSPTKPLNPRTKSKSLDPNS- 52
Db 451 NVDKRLKPLDE-----LTKSYLINSPLTSPNATIAFPNNKNNNNNNNNNNNS 506
OY 53 -----SDTYSEDOEKGEKEEKDFAQSFSDRNFLDNSIDIQOTTQH000QP000Q 106
Db 507 NNNNNNNNNNQ0000000000000000-----000000000000000Q 55
OY 107 LSFQDN-----LID-----EFSQPMSTLDTKQNPVDRYNE 142
Db 554 QOQTNNNSNNISSNNDNNNTFSPQWPSIDHDSSSSTFSPQVTTMOT---PIQFSNN 611
OY 143 NHAPTYINTSPNKSIMKKATPKASPKVAFTVTPNPELHHYEDNVEEB-----DOSOK 196
Db 611 NNNNNISMT---TVPITATATATATNKTTATATTTTTNNNNNKIQEESFGSTBNKKLND 667
OY 197 EDSVEPPLIQWOKDPQSFNDSDEDTANAVP---PPLPLHTTKPFFAOLAKN 246
Db 668 EEEINFSLSQ00WOKOLFIGNLGN---AIIKQOTPOOEOOQOIIIGTIFKKN 716

```

RESULT	ID	PRELIMINARY:	PRT:	2843 AA.
096315	Q96315			
AC	Q96315			
DT	01-DEC-2001 (Tremblrel, 19, Created)			
DT	01-DEC-2001 (Tremblrel, 19, Last sequence update)			
DT	01-JUN-2002 (Tremblrel, 21, Last annotation update)			
DE	Sgca.			
GN	Sgca.			
OS	Dictyostelium discoideum (Slime mold).			
OC	Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.			
OX	NCBI_TaxID=44683;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	PubMed=11500361.			
RA	Reolots J., Melina M., Schaap P., Van Haastert P.J.;			
RT	"The Dictyostelium homologue of mammalian soluble adenylyl cyclase			
RL	encodes a guanylyl cyclase.";			
EMBL	EMBL J. 20:4341-4348(2001).			
EMBL	AF361947; AF362097.1; -.			
DR	InterPro: IPR001054; G_Cyclase.			
DR	InterPro: IPR001440; TPR.			
DR	Pfam: PF00211; guanylate_cyc; 1.			
DR	SMART: SM00028; TPR; 3.			
DR	PROSITE: PSS0125; GUANYLATE_CYCLASES_2; 2.			
SO	SEQUENCE 2843 AA; 315362 MW; A574704219FPB12 CNC64;			

Query Match	10.28;	Score 142;	DB 5;	Length 2843;
Best Local Similarity	23.28;	Pred. No. 0.19;		
Matches 68;	Conservative 43;	Mismatches 106;	Indels 76;	Gaps 13;

[illegible]

RESULT 7			
0871A0			
ID	0871A0	PRELIMINARY;	PRT; 734 AA.
DT	01-JUN-2002 (TrEMBLrel. 21, Created)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Hypothetical 83.2 kDa protein.		
OS	Dictyostelium discoideum (Slime mold).		
OC	Eukaryota; Mycetozoa; Dictyostelidida; Dictyostelium.		
OX	NCBI_TaxID=44689;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN=AX4;		
RA	Gloechner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,		
RA	Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,		
RA	Tunggal B., Cox E., Gail M.A., Platzer M., Rosenthal A., Noegel A.A.;		
RT	"Sequence and Analysis of Chromosome 2 of Dictyostelium."		
RL	Submitted (Apr-2002) to the EMBL/Genbank/DBJ databases.		
EMBL	ACLI16980; AA008756.1; -		


```

AC Q96J09;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DR 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE K1A1816 protein (Fragment).
GN K1A1816.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 8:85-95(2001).
EMBL: AB058719; BAB47445.1; -.
FT NON_TER
FT SEQUENCE 1162 AA; 124759 MW; 0447AF72DC08877E CRC64;

Query Match
Best Local Similarity 20.9%; Score 134.5; DB 4; Length 1162;
Matches 65; Conservative 39; Mismatches 92; Indels 115; Gaps 13;

DB 4 TPQKLLPIDHSHLQLOPOSSASIFNSPTKPLNFPRTNSKPSLDPRSSSDTYTSEDOE 63
DB 455 TPQKAPRRPNCGL-LNP--AAVTAGSASGPAVVPSSDSPA---EQLKMAAQQQORA 508
OY 64 KGEKKDFAFQTSFDRNFDLNSIDIQOTIQHQQQPQQQQQLSOTDNLIDEFS---- 119
DB 509 KLMQKQK-----QHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQ 548
OY 120 -----FQTPMTSLDTLTKONPTV----- 137
DB 549 PSSPYGAFTAEKPNSEPMTPQAFNNQNPVTPPMANNLQTTNNNYLPQNHMMNINQOPN 608
OY 138 ----DKVNEHH-APVYINTSP-----NKSIMKATPK-ASPKNVAFVTNPDIHHYPNRV 187
DB 609 NLGTNSLNKQHNLTLYGNTKPLTHFNADLSQRMTPPVANPNK-----NP-LMYITQQQ 661
OY 188 EEDDOSQKEDSVPEPLLIQHQMKDPDSQFNYSDEDTNASVPP---TPPLHTTKTPFAQLN 244
DB 662 QQQQQQQQQQQQQQQQQP-----PPQLQAPRAHNSDDQKRLILL 698
OY 245 KNEVNSEPEA 255
DB 699 KQGVNMQPMA 709

RESULT 14
OQKX6 PRELIMINARY; PRT; 1171 AA.
AC Q9KX6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DR 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Bone sialoprotein-binding protein.
GN BBP.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=024;
RX MEDLINE=20115096; PubMed=10642520;
RX Tung H.S., Guss B., Hellman U., Persson L., Rubin K., Ryden C.;
RT "A bone sialoprotein-binding protein from staphylococcus aureus: a
RT member of the staphylococcal sar family 1."
RL Biochem. J. 345:611-619(2000).

```

```

DR EMBL: Y18653; CAB75732.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs: TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs: TIGR01168; YSIRK_signal; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 1171 AA; 127123 MW; C5BC812F9DA5A884 CRC64;

Query Match
Best Local Similarity 25.2%; Score 134.5; DB 2; Length 1171;
Matches 51; Conservative 30; Mismatches 80; Indels 41; Gaps 8;

OY 2 NSPFSKLL--PI-----DKHSHLQLOPOSSA-----SIFNSPTPLNFPRTNSKPSL 47
DB 81 NSTQKNDLTNPKKETMTDHSQEAKEAPTTSTQQQQNNATSTETTKQNEKNVAKPST 140
OY 48 DPNSSDITYTSEQDEKGEKKDTAFQTSFDRNFDLNSIDIQOT-----IQHQQ 98
DB 141 DKTATEDTSTVLEKKKAPNNNTNDVTTKPS-----TSEIQTPTTPTQESTNIENSQ 191
OY 99 QQPQQQQQLSOTDNNLIDEFSFQTPMTSLDTLTKONPTVDKVN-----NHAPTYINTSPNK 155
DB 192 PQTPP-----SKVDNQVDTTTPKEPVNSKEELKNP--EKLKELVNDSTDRSTKPA 245
OY 156 SIMKATPKASPKKVAFTVNP 177
DB 246 TAPTSVAPKRVNAKIRFAVAQP 267

RESULT 15
OQ9W46 PRELIMINARY; PRT; 1141 AA.
AC Q9W46;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DR 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ser-asp rich fibrinogen-binding, bone sialoprotein-binding
DE protein.
GN SDRE OR SA0521.
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Iian J., Ito T., Kanamori M.,
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Oi Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
EMBL: AP003131; BAB41752.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs: TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs: TIGR01168; YSIRK_signal; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1141 AA; 124026 MW; 445419D086C54F8 CRC64;

Query Match
Best Local Similarity 22.4%; Score 134; DB 16; Length 1141;
Matches 60; Conservative 37; Mismatches 131; Indels 40; Gaps 8;

OY 10 PIDHSHLQLOPOSSASIFNS-----PTKPLNFPRTNSKPSLDPRSSSDTYT 57
DB 91 PIKKEITNDQPEAKKESTTSTTQQQQNNVATTTETKPNIEKENVAKPSTDKTATEDTSV 150

```


This Page Blank (uspto)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:13:06 ; Search time 65.6274 seconds
(without alignments)
3378.605 Million cell updates/sec

Title: US-09-964-858-1
Perfect score: 8631
Sequence: 1 MNSTPSKLLPIDKSHLQLQ.....WVNLMLQDQDQDQDQSSSQ 1664

Scoring table: BLASTSUM62
Gap 10.0, Gapext 0.5

Archived: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8631	100.0	1664	20 AAW99462	C.albicans alpha-1
2	8631	100.0	1664	23 AAU79331	Candida albicans 1
3	8631	100.0	1664	23 AAEL9799	Candida albicans 1
4	1211	14.0	236	20 AAW99456	Amino acids 218-45
5	1211	14.0	236	23 AAEL9800	Candida albicans 1
6	382.5	4.4	3257	22 ABB67502	Drosophila melanog
7	375	4.3	10182	23 ABB38314	Staphylococcus epi
8	348.5	4.0	3696	23 ABB40235	Staphylococcus epi
9	338	3.9	5024	22 AAG82935	S. epidermidis ope
10	336	3.9	6815	22 ABB6811	Drosophila melanog

11	333.5	3.9	3111	22 ABB60327	Drosophila melanog
12	327	3.8	6281	22 AAU37403	Staphylococcus aur
13	325.5	3.8	2768	22 ABB68397	Drosophila melanog
14	314	3.6	4498	22 ABB58595	Drosophila melanog
15	311.5	3.6	1803	22 ABB65391	Drosophila melanog
16	310.5	3.6	1468	22 ABB66291	Drosophila melanog
17	309.5	3.6	2437	22 AAU34338	Staphylococcus aur
18	307.5	3.6	2519	22 ABB16636	Novel human diagno
19	303.5	3.5	1970	22 ABB64827	Drosophila melanog
20	303	3.5	2346	22 ABB65519	Drosophila melanog
21	300	3.5	4134	20 AAY31946	Plasmodium falcipa
22	298	3.5	2025	22 AAU34207	Staphylococcus aur
23	298	3.5	3158	22 AAU37018	Staphylococcus aur
24	297	3.4	2951	22 ABB60291	Drosophila melanog
25	295.5	3.4	1279	22 AAG83047	S. epidermidis ope
26	292.5	3.4	3080	22 ABB64877	Drosophila melanog
27	289.5	3.4	1852	22 ABB61170	Drosophila melanog
28	289.5	3.4	2344	22 AAU37120	Staphylococcus aur
29	287	3.3	5533	22 ABB65772	Drosophila melanog
30	287	3.3	5560	22 ABB71160	Drosophila melanog
31	286.5	3.3	1420	22 ABB63410	Drosophila melanog
32	285.5	3.3	1979	21 AAB18171	Plasmodium falcipa
33	284.5	3.3	1095	22 AAG83030	S. epidermidis ope
34	284.5	3.3	2663	22 AAM39097	Human polypeptide
35	282.5	3.3	2633	22 ABB60505	Novel human diagno
36	282.5	3.3	2688	22 AAM40883	Human polypeptide
37	281.5	3.3	3238	22 ABB71715	Drosophila melanog
38	280	3.2	2368	22 AAU34139	Staphylococcus aur
39	280	3.2	2368	22 AAU36796	Staphylococcus aur
40	279	3.2	2779	22 ABB62371	Drosophila melanog
41	276.5	3.2	1400	22 AAB83348	AAF-2 protein sequ
42	276.5	3.2	1863	18 AAB23285	Human breast and o
43	276	3.2	2194	22 AAM40114	Human polypeptide
44	274.5	3.2	1712	22 ABB60536	Drosophila melanog
45	274	3.2	1177	23 ABB40261	Staphylococcus epi

ALIGNMENTS

RESULT 1
AAW99462
ID AAW99462 standard; Protein: 1664 AA.
XX AAW99462;
AC 08-JUN-1999 (first entry)
XX
DT
DE C.albicans alpha-INTP protein.
XX
KW Integrin-like motif; vaccine; immune response; antibody; inhibition;
KW adhesion; endothelial cell; pathogenesis; infection; probe.
XX
OS Candida albicans.
XX
PN US5886151-A.
XX
PD 23-MAR-1999.
XX
PF 03-MAY-1996; 96US-0642846.
XX
PR 03-MAY-1996; 96US-0642846.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Bendel CM, Gale CA, Hostetter MK, Kendrick K, Tao NJ;
XX WPI: 1999-242618/20.
XX N-PSDB: AAX25885.
XX
PT New isolated Candida albicans protein with integrin-like motifs
XX
PS Examples: Column 13-14: 21pp; English.

XX This sequence represents the *Candida albicans* alpha-INT1 protein which
CC contains integrin-like motifs. The protein was used to derive peptides
CC AMW9456-W9461 used for producing vaccines for stimulating an immune
CC response. The antibodies can inhibit the adhesion of *C. albicans* to
CC cells, particularly endothelial cells. This blocking activity of the
CC adhesion to cells can reduce or prevent subsequent events in the
CC pathogenesis of invasive candidal infection.

SQ Sequence 1664 AA;

Query Match	100.0%	Score 8631	DB 20	Length 1664
Best Local Similarity	100.0%	Pred. NO. 0		
Matches 1664	0	Mismatches	0	Gaps 0

Qy	1	MNPSKLLPDKHSHLOLOPOSSASAIENSPKPKPLNFPNTNKSPLSLDPSSSDYTSBQ	60
Db	1	MNSTPSKLLPLDKHSHLOLOPOSSASAIENSPKPKPLNFPNTNKSPLSLDPSSSDYTSBQ	60
Qy	61	DOEGKREBKDYTAQTSFSDRNFJLDSNIDJQOTIOHQOQOPOQOQOLOSOTDNNLIDEFSEF	120
Db	61	DOEGKREBKDYTAQTSFSDRNFJLDSNIDJQOTIOHQOQOPOQOQOLOSOTDNNLIDEFSEF	120
Qy	121	QTPMTSLDLTKQNPYDYKVENHNAPLYITSPNKSIMKAPKASPKVAFTVTPNPEIH	180
Db	121	QTPMTSLDLTKQNPYDYKVENHNAPLYITSPNKSIMKAPKASPKVAFTVTPNPEIH	180
Qy	181	HYPNRAREEEDSOQKEDSVPEPLIOHQMDPOFPNYSDEDTNAAVPPPLHTTKPTFA	240
Db	181	HYPNRAREEEDSOQKEDSVPEPLIOHQMDPOFPNYSDEDTNAAVPPPLHTTKPTFA	240
Qy	241	QLLNKNNEVNSEPALTDMLKRENFNSJLSDEKVALYISPTNNNNSKVNSDMDSHLQNL	300
Db	241	QLLNKNNEVNSEPALTDMLKRENFNSJLSDEKVALYISPTNNNNSKVNSDMDSHLQNL	300
Qy	301	QDASKNTNENIHLSPALAPKNDIENPLNSTNADISLRSSGSSOSSLOSTRDNRYL	360
Db	301	QDASKNTNENIHLSPALAPKNDIENPLNSTNADISLRSSGSSOSSLOSTRDNRYL	360
Qy	361	ESVSGSKKKNPGLSLNDGJIKGFSDEVESELPRDLSROKLETTKHDAPEHNNEFTDA	420
Db	361	ESVSGSKKKNPGLSLNDGJIKGFSDEVESELPRDLSROKLETTKHDAPEHNNEFTDA	420
Qy	421	KSTNTNKGOLLVSSDHLJDSFDRSYNHTQESILNLNLSASOOSISUNALEKOROTQEOBQ	480
Db	421	KSTNTNKGOLLVSSDHLJDSFDRSYNHTQESILNLNLSASOOSISUNALEKOROTQEOBQ	480
Qy	481	TQAAPEPEEFSFQNTKVKQEPKSNLEFVAVTTIKKEPYSKTEIKAKKREFSSILRIKNE	540
Db	481	TQAAPEPEEFSFQNTKVKQEPKSNLEFVAVTTIKKEPYSKTEIKAKKREFSSILRIKNE	540
Qy	541	DETAEPADIHPKKENANSIHEVETDALLKALUNDEESPTONSTKMSIRPHIDSQMKLE	600
Db	541	DETAEPADIHPKKENANSIHEVETDALLKALUNDEESPTONSTKMSIRPHIDSQMKLE	600
Qy	601	DSNDGREDNDDISRFEKSDILNDVQSOTSDIIGDKYGNSSSEITTTKTLAPRSDNNDEXN	660
Db	601	DSNDGREDNDDISRFEKSDILNDVQSOTSDIIGDKYGNSSSEITTTKTLAPRSDNNDEXN	660
Qy	661	SKSLJEDPANNESLQOOLLEVHTKEDDSILANSSNIAPEBELTLPVYEANDYSFNDVYTT	720
Db	661	SKSLJEDPANNESLQOOLLEVHTKEDDSILANSSNIAPEBELTLPVYEANDYSFNDVYTT	720
Qy	721	FDAVSSFEESLSREHEDSKPINFISTIMHGOEKQKHQIHKVPYTKOITIASYOQYKKECS	780
Db	721	FDAVSSFEESLSREHEDSKPINFISTIMHGOEKQKHQIHKVPYTKOITIASYOQYKKECS	780
Qy	781	RVTSDKAKIENAIQKFKFKEVNVMRSRVVSPMDDDLNVSOFLPELSEDSGFKDLNPNANS	840
Db	781	RVTSDKAKIENAIQKFKFKEVNVMRSRVVSPMDDDLNVSOFLPELSEDSGFKDLNPNANS	840
Qy	841	NNTNRPSEFPPLSTKNVLSNIDNDPVVEPEPKSYAEIRMARISANKAAPQAPPLP	900
Db	841	NNTNRPSEFPPLSTKNVLSNIDNDPVVEPEPKSYAEIRMARISANKAAPQAPPLP	900

Dd	841	NTNRPNSFTPLSTKVNLSINDDPNVVERPERKSTAEIRNARRLSANKAARDNPRLPR	900
Qy	901	QROPSTFSNSNKKVSFRVPTFEIRKTSALSALRCDMUNDIFEDGAGSKPTIAEGSKJT	960
Dd	901	QROPSTFSNSNKKVSFRVPTFEIRKTSALSALRCDMUNDIFEDGAGSKPTIAEGSKJT	960
Qy	961	LPSMDKDDVKRI LAKKGGVITODEYIAAKLVDOKPKKNSIVTREDRYEELAOPTASIHNAIT	1020
Dd	961	LPSMDKDDVKRI LAKKGGVITODEYIAAKLVDOKPKKNSIVTREDRYEELAOPTASIHNAIT	1020
Qy	1021	IDSSISYGRPDSISNDMLRYULSDELKKRPYLLSADRLPFMEOEVRNPLRYSVLYNHGASAA	1080
Dd	1021	IDSSISYGRPDSISNDMLRYULSDELKKRPYLLSADRLPFMEOEVRNPLRYSVLYNHGASAA	1080
Qy	1081	TNSSMLPERDELINSAPRVNNSNDVVALISGNASTISFNOLDIMNDDOATIGOKIQORP	1140
Dd	1081	TNSSMLPERDELINSAPRVNNSNDVVALISGNASTISFNOLDIMNDDOATIGOKIQORP	1140
Qy	1141	ASKSANTVKGDDDLGASARETPRTPTKKESISSKPAKLSASAPRSPKIGSPVAVIKKN	1200
Dd	1141	ASKSANTVKGDDDLGASARETPRTPTKKESISSKPAKLSASAPRSPKIGSPVAVIKKN	1200
Qy	1201	GSINGIERPKATPKPKKSPGNGEISNNHYKROGGSIPSSGSEHOONPMSVSPQUTYDA	1260
Dd	1201	GSINGIERPKATPKPKKSPGNGEISNNHYKROGGSIPSSGSEHOONPMSVSPQUTYDA	1260
Qy	1261	TSTYARDENKDVONKPREKQKQKNNHNNNNNNNNKOKTDIPGVVDDDELPRDGLDERGKLPRF	1320
Dd	1261	TSTYARDENKDVONKPREKQKQKNNHNNNNNNNNKOKTDIPGVVDDDELPRDGLDERGKLPRF	1320
Qy	1321	VLGIKTNLPDINTNGKRPYLLTDNGVNCVUTPREUMMDHNAVALGKERELTVADSLERFL	1380
Dd	1321	VLGIKTNLPDINTNGKRPYLLTDNGVNCVUTPREUMMDHNAVALGKERELTVADSLERFL	1380
Qy	1381	TLKASYSKRGTELEVETKEKVVUKSNNRPLSLRESKDIITTTGFVREYKDTMANKERARDG	1440
Dd	1381	TLKASYSKRGTELEVETKEKVVUKSNNRPLSLRESKDIITTTGFVREYKDTMANKERARDG	1440
Qy	1441	SFACASYIDLQOFEOQLTGKASOFDLNCFMEWETMSNGNOPMKRGKRYUKIAOLEVMILVLP	1500
Dd	1441	SFACASYIDLQOFEOQLTGKASOFDLNCFMEWETMSNGNOPMKRGKRYUKIAOLEVMILVLP	1500
Qy	1501	RSDPRELLPYSINSAYESINELNNEOONNYFEGLHOGGDCPIFKKRFKELMGTSLLAHS	1560
Dd	1501	RSDPRELLPYSINSAYESINELNNEOONNYFEGLHOGGDCPIFKKRFKELMGTSLLAHS	1560
Qy	1561	EISHKRTAKINLSKVVDYLIVYDKENIDRSNNHNFSSVULLDPAFKTKRANGELIDFCARN	1620
Dd	1561	EISHKRTAKINLSKVVDYLIVYDKENIDRSNNHNFSSVULLDPAFKTKRANGELIDFCARN	1620
Qy	1621	KHEKKIMTONLOEITTYENFRFRPQWNLMLQOOQOOQOOQSSSQ 1664	
Dd	1621	KHEKKIMTONLOEITTYENFRFRPQWNLMLQOOQOOQOOQSSSQ 1664	
RESULT 2			
AAU79331			
AC	AAU79331	standard: Protein: 1664 AA.	
XX	AAU79331:		
XX	02-JUL-2002	(first entry)	
DE	Candida albicans integrin-like protein, Intlp propertide.		
XX	Integrin-like protein; Intlp propertide; bactericide; vaccine;		
KW	yeast infection; thrush; paronychia; candidal vaginitis;		
KW	immunocompromised host; cancer patient; transplant patient;		
KW	premature newborn; human immunodeficiency virus infection;		
KW	HIV infection; periodontitis; oral ulceration; esophagitis.		
XX	Candida albicans.		
DS			


```

QY 841 NNTNRPSFTPLSTKKNVLSNIDNDPNVVEPPPKSYAEIRNARLSANKAAPNOAPPLPP 900
Db 841 NNTNRPSFTPLSTKKNVLSNIDNDPNVVEPPPKSYAEIRNARLSANKAAPNOAPPLPP 900
QY 901 QROPSTSNSSNKKRVSRFRVPFTEIRRTSSALAPCDMYNDIFDDGAGSKPTIKAEGMKT 960
Db 901 QROPSTSNSSNKKRVSRFRVPFTEIRRTSSALAPCDMYNDIFDDGAGSKPTIKAEGMKT 960
QY 961 LPSMKDDVKKRLNAKKGVTQDEYINAKLYDOKPKKNSIVTDPEDRYEELQOTASTIHMT 1020
Db 961 LPSMKDDVKKRLNAKKGVTQDEYINAKLYDOKPKKNSIVTDPEDRYEELQOTASTIHMT 1020
QY 1021 IDSSLYGRPDSISTDMLPLSLDELKKPPALLSADRLFEQOEHPHRSNSVYVHPGAGAA 1080
Db 1021 IDSSLYGRPDSISTDMLPLSLDELKKPPALLSADRLFEQOEHPHRSNSVYVHPGAGAA 1080
QY 1081 TNSMKLPEPDFELINSPARNVSNNSDNVAISGNASTISFNQIDMNFDDGATTGOKIQEOP 1140
Db 1081 TNSMKLPEPDFELINSPARNVSNNSDNVAISGNASTISFNQIDMNFDDGATTGOKIQEOP 1140
QY 1141 ASKSANTVRGDDDLASAPERTPTPKKESISSKPAKISSASPRKSPKIGSPVRVIRKN 1200
Db 1141 ASKSANTVRGDDDLASAPERTPTPKKESISSKPAKISSASPRKSPKIGSPVRVIRKN 1200
QY 1201 GSIAIEPIPKATNPKPKSFOGNEISNHNKVRDGGISPSGSEHOCHNPMSVSPQYTD 1260
Db 1201 GSIAIEPIPKATNPKPKSFOGNEISNHNKVRDGGISPSGSEHOCHNPMSVSPQYTD 1260
QY 1261 TSTVDENKDVOKHREKOKHNNHNNHNNHKKOKTDIPGVVDEIDPVGLOERKLFPR 1320
Db 1261 TSTVDENKDVOKHREKOKHNNHNNHNNHKKOKTDIPGVVDEIDPVGLOERKLFPR 1320
QY 1321 VLGITNINLPDINTHKGRTLLDNGVCHVTTPRYMDNDHNAIGEFELTVADSLFTL 1380
Db 1321 VLGITNINLPDINTHKGRTLLDNGVCHVTTPRYMDNDHNAIGEFELTVADSLFTL 1380
QY 1381 TLKASAYERPGTLVEYEKKVYSRRRLSRLFGSKDITTTKFPVPEVADTMANKFAPDG 1440
Db 1381 TLKASAYERPGTLVEYEKKVYSRRRLSRLFGSKDITTTKFPVPEVADTMANKFAPDG 1440
QY 1441 SFARCYIDLOQFEDDITGKASOFDLNCFNEMETMSNGNOPMKRGKYKIAOLEVKMLYVP 1500
Db 1441 SFARCYIDLOQFEDDITGKASOFDLNCFNEMETMSNGNOPMKRGKYKIAOLEVKMLYVP 1500
QY 1501 RSDPREILPTSRSAVESINEINNEONNFFEGYLHOEGGDCPIFKKRFKLMGTSLAHS 1560
Db 1501 RSDPREILPTSRSAVESINEINNEONNFFEGYLHOEGGDCPIFKKRFKLMGTSLAHS 1560
QY 1561 EISHKTRAKINLSKVVDLIYVDKENIDRSNHRNFSVDLLDHAFAIKFANGELIDFCAPN 1620
Db 1561 EISHKTRAKINLSKVVDLIYVDKENIDRSNHRNFSVDLLDHAFAIKFANGELIDFCAPN 1620
QY 1621 KHEMKIWIIONLOEIIYRNFRQPVVNLMLQOQOQOQOQOQSSQ 1664
Db 1621 KHEMKIWIIONLOEIIYRNFRQPVVNLMLQOQOQOQOQOQSSQ 1664

```

RESULT 4

AAW99456
ID AAW99456 standard; protein; 236 AA.

AC AAW99456;

DT 08-JUN-1999 (first entry)

DE Amino acids 218-453 of C.albicans integrin-like protein 1.

KW Integrin-like motif; vaccine; immune response; antibody; inhibition;
adhesion; endothelial cell; pathogenesis; infection.

OS Candida albicans.
XX

```

PN US5886151-A.
XX 23-MAR-1999.
XX
PF 03-MAY-1996; 96US-0642846.
XX
PR 03-MAY-1996; 96US-0642846.
XX
PA (MIND ) UNIV MINNESOTA.
XX
PI Bendel CM, Gale CA, Hostetter MK, Kendrick K, Tao NJ;
XX WPI: 1999-242618/20.
XX
DR New isolated Candida albicans protein with integrin-like motifs
XX
PS Claim 1: Column 35; 21pp; English.
XX
CC Peptides AAW99456-w99461 are derived from a Candida albicans protein
CC with integrin-like motifs, alpha-INP1. This sequence represents amino
CC acids 218-453 of alpha-INP1. The peptides can be used for producing
CC vaccines for stimulating an immune response. The antibodies can inhibit
CC the adhesion of C.albicans to cells, particularly endothelial cells.
CC This blocking activity of the adhesion to cells can reduce or prevent
CC subsequent events in the pathogenesis of invasive candidal infection.
XX
SQ Sequence 236 AA:

```

Query Match 14.0%; Score 1211; DB 20; Length 236;
Best Local Similarity 100.0%; Pred. No. 5.2e-67;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 218 SDEDTNNAVPPPLHTHTPTFAOLLKKNNEVSEPEALTDMLKRENFSLSLDEKYNL 277
Db 1 SDEDTNNAVPPPLHTHTPTFAOLLKKNNEVSEPEALTDMLKRENFSLSLDEKYNL 60
QY 278 YLSPTNNNSKNVSDMDSHLQNDASKNKTENINHLSPALKAAPKNDIENPLNSLTNAD 337
Db 61 YLSPTNNNSKNVSDMDSHLQNDASKNKTENINHLSPALKAAPKNDIENPLNSLTNAD 120
QY 338 ISLRSSGSSQSLQSLRNDNRVLESYVGPSPKKNVPGSLNDGKIGFSDEYVESLPRDLS 397
Db 121 ISLRSSGSSQSLQSLRNDNRVLESYVGPSPKKNVPGSLNDGKIGFSDEYVESLPRDLS 180
QY 398 RDKLETTKEHADAEHNENFTDAKSTNTNGQLVSSDOHLDSPDRSYNHTEQSIL 453
Db 181 RDKLETTKEHADAEHNENFTDAKSTNTNGQLVSSDOHLDSPDRSYNHTEQSIL 236

```

RESULT 5

AAE19800
ID AAE19800 standard; protein; 236 AA.

AC AAE19800;

DT 18-JUN-2002 (first entry)

DE Candida albicans integrin-like protein (alphaIntlp) fragment.

KW Integrin-like protein; alphaIntlp; gene; fungicide; contraceptive;
HIV; human immunodeficiency virus; vaccine; cholera.

OS Candida albicans.

PN US6346411-B1.

PD 12-FEB-2002.

PF 08-MAR-1999; 99US-0264604.

PR 03-MAY-1996; 96US-0642846.

PA (MIND) UNIV MINNESOTA.

XX Hostetter MK, Gale CA, Bendel CM, Tao N;
XX WPI: 2002-224995/28.
XX
XX Polynucleotide encoding *Candida albicans* protein with integrin-like
XX motifs and protein, useful as vaccines and for raising antibodies for
XX inhibiting adhesion of pathogen to cells preferably epithelial cells
XX
XX Example: Column 17-18; 25pp; English.
XX
XX The present invention relates to an isolated polynucleotide encoding
XX *Candida albicans* protein with integrin-like motifs. *Candida albicans*
XX protein is utilized in vaccines (for gastrointestinal pathogens like
XX cholera) and as antigens to prepare anti-peptide antibodies, which are
XX utilized in inhibiting adhesion of *C. albicans* to cells preferably
XX epithelial cells. In the genitourinary tract, expression of spermicides
XX by *S. cerevisiae* transformed with the *C. albicans* integrin-like gene on
XX an extrachromosomal plasmid could provide a cheap and infrequent method
XX of contraception. Also, synthesis of protein-based antiretroviral
XX agents could help to reduce transmission of human immunodeficiency virus
XX (HIV) in the birth canal. The present sequence is *Candida albicans*
XX integrin-like protein (alphaInIip) fragment.

SQ Sequence 236 AA:

Query Match 14.0%; Score 1211; DB 23; Length 236;
Best Local Similarity 100.0%; Pred. No. 5,2e-67;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SDEDTNASTVPPTPLHTTPTFAQLLNKNNVSEPEALTDKMKRENSNLSDEKYNL 277
DB 1 SDEDTNASTVPPTPLHTTPTFAQLLNKNNVSEPEALTDKMKRENSNLSDEKYNL 60
QY 278 YLSPPTNNNNKSNVSDMDSHLQNLQDASKNKTNNENHNLSPALKAPKNDIENPLNSLTNAD 337
DB 61 YLSPPTNNNNKSNVSDMDSHLQNLQDASKNKTNNENHNLSPALKAPKNDIENPLNSLTNAD 120
QY 338 ISLRSSGSSSSLOSLRNRRVLESVPGSPKVPNGLSINDGKGGSDVESLPRDLS 397
DB 121 ISLRSSGSSSSLOSLRNRRVLESVPGSPKVPNGLSINDGKGGSDVESLPRDLS 180
QY 398 RDKLETTKEHADPEHNNENFIDAKSTNTNKKGOLLVSSDDHLSFDRSYNHTQESIL 453
DB 181 RDKLETTKEHADPEHNNENFIDAKSTNTNKKGOLLVSSDDHLSFDRSYNHTQESIL 236

RESULT 6
ABB67502
ID ABB67502 standard; Protein; 3257 AA.

ABB67502;

DT 26-MAR-2002 (first entry)
DE *Drosophila melanogaster* polypeptide SEQ ID NO 29298.
XX
XX *Drosophila*: developmental biology; cell signalling; insecticide;
XX pharmaceutical.
OS *Drosophila melanogaster*.
PN WO200171042-A2.
XX
XX 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PERE) PE CORP NY.
XX

PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
DR N-PSDB; ABL1605.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from *Drosophila* and for elucidating cell signalling and cell-cell
XX interactions
XX
XX Disclosure; SEQ ID NO 29298; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL6175) and the encoded proteins
CC (AB57737-AB872072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 3257 AA:

Query Match 4.4%; Score 382.5; DB 22; Length 3257;
Best Local Similarity 19.6%; Pred. No. 3,6e-14;
Matches 373; Conservative 292; Mismatches 679; Indels 559; Gaps 88;

QY 4 TFSKLLPID-----KSHLLOPOSSSASIFNSPTKP-----INFPTNSKPSLDPS 51
DB 588 TGSNTNLPQDDIMASHNQTDLKCAPDRVALDKSESTPVEEOLCKYDTPSDTALDESK 647
QY 52 SSDT---YTSDDQEKKEE-KKDTAFQTSFDRNFDLNSIDIOQTIOHQOQOPOQOQO- 106
DB 648 VESAKNHTELEDDKDKDKETQKESPNGSKETN---ENSVITVEVELPAKKAKEAKEA 704
QY 107 ---LSQTDNNLLDSEF-----QTPMTSTDLTLQONPTVDKVNENHAPT 147
DB 705 GNIVESDSQLEDFRLAEELIAEYGVKEANEVSVTSVG-RCNPVETETVKELEDET 762
QY 148 YINTSP-----NKSIRKATPKASPKKVAFTYTNELIHHYPDNRVEEDQOQOQED 198
DB 763 ISEVPAQNDQSSVEQDTLADKENVEKSPYKAKSSS-----KDEPAEENLAPPDQ 816
QY 199 SVE---PPLIOHQWD-----PSQFNYSDED-TNASTVPPTPLHTTPTFAQLLNKNE 248
DB 817 PLEOQKTPVAKNQDQKHEHNEAPKAESLSVSDIPSSVPSKRRHNSP--ANTPKKSKE 874
QY 249 V-----NSPEALTDKMKLKRENFNSLSL--DEKV-----NLYLSTPTNNN 286
DB 875 IEALQSSVPRRALRSQKATPQMLRESRSKRTLTETLLMDPTMRSSPRIGRSPAESHS 934
QY 287 SKNVSDMS-----HLQNLQDASKN-----TNE 310
DB 935 SHERSPMEKKYVSKLADLTIIDKEKTELKSLDASSTKTKYKTTKTTASDTSLTDE 994
QY 311 NTHNLSFALKAPKNDIENPLNSLTGNADISLRSSGSSSSLOSLRNDNRVLESVPG--SPK 368
DB 995 NPSSSKTEMKKLLG---KPLK-----AKKMSRSTSEVKAIDSN--EDIPSTFISK 1042
QY 369 KVPGLSLNDGKIGFSDEYVESLPR---DLSRDKLETTKEHADPEHNNENFIDAKSTNT 425
DB 1043 CVEEHLTSSSEBO--KDEKEELLCRPOIDCTNTLE--OSTALETPTQ--VEEKRSNR 1096
QY 426 NKGOLLVSS-----DDHLSFDRSYNHTQESILNLSNASQOSISLNALEQR--- 473
DB 1097 RKSRIENKEFTETDITSLDIDAKKAENASLEISMRCJKTLETQSDPVTAKKRNRSGR 1156
QY 474 -OTQOEOTQAAPPEETS---FSDNIKVKQEPKSNLEFVKVYTIKKEPVASATEIKAPRE 529
DB 1157 LSRKEKSVINAKSKSKSPSAISOSTERKOLLNEN-----PSKKD--KTEGSGNKKE 1208

Qy	530	FSHILIRIKNEDELAPADIRHKKEHEA-NSHVEDTDALIKKALNDEDESDTTONSKMS	588
Db	1209	---AVGPELDKTESSSTNIIIDKSSNEPSDMSQPSD---RLNOKSAPFTKLSSISSP	1260
Qy	589	IRFHIDSDMKL-----DSDNDGREINDDISPREKEDI-----LNDV	625
Db	1261	KKIMKODDKLDAISKGGSNPITRTGDSQOTQTKKQENDTKHEEDSSKLKANIDET	1320
Qy	626	SQTSIDIIGDKYGNSSSEITTK-TLAPRSDNNDKENSKSLED-----PANNESL	673
Db	1321	KSSSEKDAEPISKSDSODSAKPRLSPKSRNKKRKKNEKKPNDISIAESDIEGQVNETYV	1380
Qy	674	QOOLEVP-HIKEDDSILANSSNAP-----PELTLPVVEANDY	711
Db	1381	QATCSPTSESENKKDMKWSDETNEEPLSTETIGIRIRKGOAFHIENPMD-DLHITPONEN	1439
Qy	712	SSFNVDYKRTPDAYSPEESLSREHENDSKPINFISIMHKQOKKHQIHKVPTKOTIASY	771
Db	1440	OSIAGV--NEKOVPLPESY-----ESD-----PLMKLPTTYLMCT	1475
Qy	772	QOYKNEOESHVTSIDKYLIPALIOFKFEKEVNVMSRRVYSPMDIDLNVSOQLPELSEDSGF	831
Db	1476	KNKSLLSASBEDDIYLER-----QKLITTSKGDSNPOLDNAN-----NLEHSST	1520
Qy	832	KDLNFANYSNNKTRPNSEFTPLSTKNVLSIINDPVNERPEP-KSYAIEIRARLSANKA	890
Db	1521	QDPKEHEFSQOT--FT--DMSDIIIPCSCTKSOIIVPEPTTPTKSSDQTKN-----SFI	1567
Qy	891	APNAPPLP-----POQOPSTGSSNKNKVRNRVYTFE-----IRTS	929
Db	1568	TPNNSPSPKSRNVSKSKEAKRLDNSFEESONASASASKVOKELRTPTJASCRLRYLIKTRP	1627
Qy	930	SALAPCMYNDIEDDFEGAGSKPTIK-AEGMKTLPSMD-----KDDYKRI	972
Db	1628	TSLPLTMSRKSIFKKTIPAKKRLTLKLEMEKTPSEPSVSLGEVNPDSQVYAESAVAL	1687
Qy	973	LNAAKGYQODEYINAKLVODKPKKNSIYVD-----PEDRYE-ELQO--TASHINATIDSSI	1025
Db	1688	HESDRDLESNEIEPEEFEDTEPASAEDTDNKLKKEDHELEVUDICAAKNPITDOST	1747
Qy	1026	YGRDJSIST--DMLPYLSDLEKPKPALLASD-----RLRMOEVPPLNS-----	1068
Db	1748	KDASNSKSTSDVLYQETKDLIS--NSLINAQGEDTPIKELTEEBEVRNNKTVINDEDESKQ	1804
Qy	1069	-----NSVLVHPGAGATNS-----SMLEP-----DFELINSFAR	1099
Db	1805	EILKDLPERDAALEEDPTASAKAAEEMDLIKEKSNVKSVALPERDVTDDBELAQSPTR	1864
Qy	1100	NVSNNSDNLVAINSGNASTISF-----NOLDNPFDOAT-----IGOKIDQEP	1140
Db	1865	N-SETTISVYDDPERPSTSVYKSLRLKREADSOPRDEAKRKORODVEKSLGKKEQYK	1923
Qy	1141	A-----SKSANTVGR-----DDGLASAPETPRTPTKKESISSK	1174
Db	1924	ARRQOLAEVEERPSLKSSTESKSTSVQOKYISIIIGNETIWSSTAPRTRETNREASTS	1993
Qy	1175	PAKLUSA-----SPKRSPTKIGSPRYVLKKGSIAGIEPTPKAT-----HK	1215
Db	1984	PSAKSNAVQEAKEHNHETTKHIIIDRPPCKKLILHDSRPAEKK-----PMVOTLLSSTLSQK	2039
Qy	1216	PKKSPQGNELSNHVR-----DGLISPSGSEHQOONDSMWSVPSQYTDATYST	1263
Db	2040	PSLIDDCSPL--KIRSLKLSIADENIDDOGSIFSSSVLKKNTSYVA-PRKVIISVSL	2095
Qy	1264	VPDENKQV-----HKPRKQOKQKNNNNNNNNNNNNNNKQATDIDRGVUDDIIP	1307
Db	2096	L--OSKDOVETAASSETPLTKKEKELQOKTKPRKESNKTESKKSLYOG-----P	2147
Qy	1308	DVGLQERKLEFFRYLIGIKINIL-LPDITNGHGRETLIDLGSVLCVUTTPREYMNMDHNVALIG	1366
Db	2148	QMKQOKSE---ANSGRKILNKYLAKSTESKRTVETVYG-----KQOIGQ	2190
Qy	1367	EFELTVADSLFEIILTLKASYEKPRGTLVETEKVVVKSNNRLSRLEGSKDIIITTTKFEVPT	1426

Dd	2191	-----LEVL--KKEESKRSESLVEALSRKKQSOYQRHLSKIDGKSGESTISLPDPV	2239
Qy	1427	EVKDITWANKFAPDDGSFARCYIDLQOFEDOITGKASQFDLNCENEMETJNSNGNQPKRGKP	1486
Dd	2240	SKSEDTALKAAALP-----KETEPFYQ-DAELEKMSKGHGHNAYKN	2278
Qy	1487	YKIAOLEVKKMLVPRSDPRELLPTSIRSAV-ESINELNNEONN	1528
Dd	2279	TKTEQ-----PKSKPK---TEVRSLQAFAATFLDMSDS	2309
 RESULT 7			
ABP38314	ID	ABP38314 standard; Protein; 10182 AA.	
XX	AC	ABP38314:	
XX	AD		
XX	AE		
XX	AF		
XX	AG		
XX	AH		
XX	AI		
XX	AJ		
XX	AK		
XX	AL		
XX	AM		
XX	AN		
XX	AO		
XX	AP		
XX	AQ		
XX	AR		
XX	AS		
XX	AT		
XX	AU		
XX	AV		
XX	AW		
XX	AX		
XX	AY		
XX	AZ		
XX	BA		
XX	BB		
XX	BC		
XX	BD		
XX	BE		
XX	BF		
XX	BG		
XX	BH		
XX	BI		
XX	BJ		
XX	BK		
XX	BL		
XX	BM		
XX	BN		
XX	BO		
XX	BP		
XX	BQ		
XX	BR		
XX	BS		
XX	BT		
XX	BU		
XX	BV		
XX	BW		
XX	BX		
XX	BY		
XX	BZ		
XX	CA		
XX	CB		
XX	CC		
XX	CD		
XX	CE		
XX	CF		
XX	CG		
XX	CH		
XX	CI		
XX	CJ		
XX	CK		
XX	CL		
XX	CM		
XX	CN		
XX	CO		
XX	CP		
XX	CQ		
XX	CR		
XX	CS		
XX	CT		
XX	CU		
XX	CV		
XX	CW		
XX	CX		
XX	CY		
XX	CZ		
XX	DA		
XX	DB		
XX	DC		
XX	DD		
XX	DE		
XX	DF		
XX	DG		
XX	DH		
XX	DI		
XX	DJ		
XX	DK		
XX	DL		
XX	DM		
XX	DN		
XX	DO		
XX	DP		
XX	DQ		
XX	DR		
XX	DS		
XX	DT		
XX	DU		
XX	DV		
XX	DW		
XX	DX		
XX	DY		
XX	DZ		
XX	EA		
XX	EB		
XX	EC		
XX	ED		
XX	EE		
XX	EF		
XX	EG		
XX	EH		
XX	EI		
XX	EJ		
XX	EK		
XX	EL		
XX	EM		
XX	EN		
XX	EO		
XX	EP		
XX	EQ		
XX	ER		
XX	ES		
XX	ET		
XX	EU		
XX	EV		
XX	EW		
XX	EX		
XX	EY		
XX	EZ		
XX	FA		
XX	FB		
XX	FC		
XX	FD		
XX	FE		
XX	FF		
XX	FG		
XX	FH		
XX	FI		
XX	FJ		
XX	FK		
XX	FL		
XX	FM		
XX	FN		
XX	FO		
XX	FP		
XX	FQ		
XX	FR		
XX	FS		
XX	FT		
XX	FU		

QY	135	----	-PTVQVKNHNAHPYTIINTSPKMSIMKATKTPASAKVAFTVNPDEIHHPDNEV	188
Db	3543	KAEQVIOANAKIIEENQPSVQYUDESKEVEQALSTLNNAKASALALADQOEIQAYNOLIQ	3602	
QY	189	EEDSOQOEKEDSVEPPILOHOMKDPQFNYSDIEDTASVPPPELHTTKPTPAQLLNKNE	248	
Db	3603	PTDLNNKRPASITAYANORY-----QESNELNSTKTTWDRILKEQNPSPADVYNALNK	3655	
QY	249	VNSEPALITD-----MKIKRENFNSLSDKRYNLYLS--PTNNNSKNKNSD-----	292	
Db	3656	VREYQOKLNEBRALLQKEDNSALVPAKQEQLOQADQVPESTEGMTQITKDYNSKQAAQ	3715	
QY	293	-----MDSHLOLQDASKNKYT-----ENHILNLSFALKAPNDIENPLNSTLT-NAD	337	
Db	3716	QEISKAGQVINDGATQOISNAKTVBEALFALNNNAKGLBRADKEEIQMAYNOLITQID	3775	
QY	338	-----ISLSSGSSQSSLIQS-----LRNDNFVLESYBSPKKVNPGLSLNDG	379	
Db	3776	TSGCTPASIRKRYNEAKRSIOTQIDASKNEANSILITNDNPQVSQYTAALNKI-----	3826	
QY	380	IKGSDSEVSESLIPROLSROKLETTKEHAPHEHNENFIDAK-----	421	
Db	3827	-----KAVQBELDKAIAIMLK-----MNE-----NNNALVQAQOQLOQIVNEVDPYTGMTT	3871	
QY	422	STNTNKQOLLVSPDDHLDSDRSYNH-----TEOSIILNLSASQSOISLNALEKORQOEQ	478	
Db	3872	DTANNNYKSKKREADEIQKAOQIINNNGDTEQOITNETRWQO-----ALNALINKAKNDRA	3928	
QY	479	EQTOAAREEETSIDNIKAKQEPKSNLEFVYKTIKEVEVSTETIKAPKREFFSRILRIK	538	
Db	3929	DKSOL-----ENAYNOLIQ-----NVD-----TNGKKPASIQOQAQARAQIETQYNNAK	3972	
QY	539	NE-----DEIAEPAD-----IHPKKENANSYHVEDTALLKKALND	574	
Db	3973	SEAOIILENSPSPNEVQAOLQKVEAVOLKVNDALHILOKKNNSALVTAKNOLOQSYND	4032	
QY	575	D-----EESDTQONSTKMSIRFIHIDSQWLKEDSNDGDREDNDISRFEXS-DIL	622	
Db	4033	QPLTGTQOTDSINNYEAKRREAOASAIR---NAEAVINNCGDAPAKQISDEKSKVEQALHL	4089	
QY	623	NDVSO-----TSDIIG-----DKYGSSEITTKTILAPRSNDNKEMSKLE-----	665	
Db	4090	NDAKOQLTADTTELQTAQVQOLNRRGDTNNK-----KRSITANVKNKIQISLEQITQISAK	4142	
QY	666	DPAN-----NESILOQ-----OLEVPHRKEDDSILANSNTJAPPEELTLPVPEAND	710	
Db	4143	DNAAAVQKPIRTVQEVNNALQOVYNLOLITFALINQLOPLSNNDA-----LKAARLNL	4196	
QY	711	YSSFNDVTKT-----FDAYSFEESLSREH-----ETD-----	738	
Db	4197	ENKINQIVQJQDGMTQOSIEAYQNAKRVQAQNESNTALALINNCGADEQOITTETRVANOQT	4256	
QY	739	---SKPINFISIMHIQO---EKQK---KHQIHKVP-----TKOILASYQO---YKNE-----	777	
Db	4257	TNLIQALINGLTV--NKEPLETAKTALONNDIDQVPESTDGMTQOSVANYOKLOJANKEINTI	4315	
QY	778	-----QESVTSQKVKIPAIQO-----EFKKF-----	799	
Db	4316	NNVLANNPDVNAIKTKAEKERISNDLTOAKKNLQVDTQPELEKIRQLODELDTGNTDG	4375	
QY	800	-----EYVNASRRVSPMDDL-----NVSQFLPEL-----	825	
Db	4376	MTQOSVDNYNDLSALITIEKGKYNKLLKR--NPTEQVQESVANAQOYIQIDQANRATSLV	4433	
QY	826	-----SEDSGKQLNLFANYSNNTNRPRPSFTPLSTKVLNLSIINDP	865	
Db	4434	PDKTOLOEAKNRLENSTINOQTDJTGMTQOSLNNYNDKLAKAQO-----NLEKISKVLOGP	4489	
QY	866	NVPEPPEKSVAEIR--NARLSLANKAAPQA--PPLPEQOPQPSSTRSNSNKRKVRSPRYTF	923	
Db	4490	TV-----AEIRONTDEAANAHHKQALDTASQSLTLNREPIINHINNESHLNNAQKDNF	4540	

QY	924	EIRRTSSALRP-----	-----CDMWNDIEDDGGAGSGPTLKAEGSKTPRS--MDKDYKR	971
Db	4541	KAQVNS---APNHHTLETITKNKADLTN---	QSMFALSESIALDYENQKOENLIDLSNNKR	4594
QY	972	-----ILNAKKGV---TODEYINAKLVDOKR--	KNSIVTPREDRYEELQOTASINAT	1020
Db	4595	QDYDAVNAAKGILNQTOSPTMSADVIDQKAEDVKKRTKALDGNORLEAKQAOALNHNLT		4654
QY	1021	IDSSIYGRPDSITDMLPYLSEDELKRPYALYSADRLFMEQEVNPLRSNYSVLVHRGAGA	1080	
Db	4655	LND-----	LNDAROTLTDTIN-----HSPNINSVQAQKEKANT	4688
QY	1081	TNNSM-----LRRPDEL-----INS--PARVNSNSDVAISGNASTISENOLDNFD	1127	
Db	4689	VNTAMTOLKQTIANYUDELDHGVNIADKDKKDAYNNAAVN-----NAKL--INSDANDA	4742	
QY	1128	---DQATIGQIKQEDRPAAKSANYUGRD-----	DGLASRPETRPPTKSESIS	1172
Db	4743	QLDRLEINKTQYQRYNVTK--NDINGKDLALAEAKRDANTTIDGTYLNEQYRKA--KENVG	4799	
QY	1173	SKPAAKLSASBPARKPKIGSGPVLVKKGNSIAGIEPRTAKTRPKKSFQAGNISNKKYRD	1232	
Db	4800	KASRTNITISOLDYQNLNIAQALRS-----	YDVNNVTKANS	4838
QY	1233	GGISPPSGSE-----	HOONPSW--SVPSQYTTDATSTV-----	1264
Db	4839	NYIENEDGPKRAYQAQATVNAQTLLINQSPREMSRDVVNQKQATVNAHONLNGOQKLEQA	4898	
QY	1265	-PRNKIVQONKPRKQKJKNHNNHNNHNNKQKTDIP-----	GVYDEETPDVG	1310
Db	4899	QSSANTIEGILNPLNTQKAKKEKELVNSQTRFEVQDOLQAKSLDSSNGTAKLSIYAKOR	4958	
QY	1311	LOERKGLFEFRVLGKININLRDITNKGRTLL-----	TLDSGVNHCVTTPRE	1354
Db	4959	TVQKTSYUI-----NEDQPEQASUANDSITMGTIINKTADVDLTKTLVDNAISNISTRE	5012	
QY	1355	YNMDDHVAIAGKEBELVA--DSLERLITLKASYEKPRGTIVEV-----	TEKKVYKASRN	1406
Db	5013	N-----ALHGEOKLTATTEINAMLTPL-ADLNTPQKEAKITGALNTANHTTDTVAEQS	5064	
QY	1407	RLSRLEFGS---KDIITTKFVRTEWKDTWANKFARDGSPARCYSIID-LOOFEDQITGKAS	1461	
Db	5065	KANQINSAMHTLRQNISDSNESHVNE-----SNYINAREKONAFETALNNAKKEIYNEQQA	5119	
QY	1462	QFDLNCNE-----WETNSGNOZMKKGRPKYTKIAQLEVKMLIYVRSDRPRIETSTIRS	1514	
Db	5120	TLDSNINSINQQAQALITLTNNALDEEDLRRAK-----	ENDADEI-----	5157
QY	1515	AYESINELNEDDUNNYFEGYUHQEGGCRPIFK-----KRFKILMGTSILANSHESISKTRAK	1569	
Db	5158	-NLNLQTLDAQRNSEKGLNNSQSTTEVASOLAKAKELKMYA-----	EDLNHLINGK	5208
QY	1570	---INLSKVVDLIYVDKENIDRSNHRNFSVDLLLDHAFIKFRANGEL	1613	
Db	5209	NOMINSSKFI-----NEDANQOQAYSAIASAALKNKSNQNPFL	5247	
RESULT: B				
ABP40235				
ID ABP40235 standard; Protein; 3696 AA.				
XX	AC	ABP40235;		
XX	DT	24-JUL-2002 (first entry)		
DE	XX	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5080.		
KW	XX	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;		
XX	XX	antibacterial; gene therapy.		
OS	XX	Staphylococcus epidermidis.		
XX	XX	US6380370-B1.		

XX 30-APR-2002.
XX PD 13-AUG-1998; 98US-0134001.
XX PR 14-AUG-1997; 97US-055779P.
XX PR 08-NOV-1997; 97US-064964P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX WPI: 2002-381255/41.
XX DR N-PSDB; ABN92780.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure: SEQ ID 5080; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences
XX can also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life
XX cycle or inhibit S. epidermidis infection.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO web site.
XX
XX Sequence 3696 AA;
XX
XX Query Match 4.0%; Score 348.5; DB 23; Length 3696;
XX Best Local Similarity 18.2%; Pred. No. 5.5e-12;
XX Matches 363; Conservative 327; Mismatches 759; Indels 549; Gaps 90;
XX
XX 1 MNSTPSKLLPDKHSHLQPOSSASISTNSPTKPLNFRRTSKPSLSDNSSDRTTSHQ 60
XX DB 1187 LNSTHAIQD-EKQDALRLTQAKETAL-----NDINQATQMONVDTALTSGIONIQNTQ 1240
XX
XX 61 DOEKGEEKKDTAFQTSFDRNFDLNSIDIQOTIQHOQOQPOQOQLOLSDTNLLIDEFSE 120
XX DB 1241 VVNRKKQAKATTI-----NDI-VQOHKQSIQNNDDATTEKEEVANNLVN--AS 1285
XX
XX 121 QTPMSTLDTLKQNTFYDVKNENHAPTYNTSPKNSIMKATP-----KSPKVAFTVTN 176
XX DB 1286 QONVTSKIDNATTTNMOIGIVSDGROSIMATTPDTSIKRNKANDIDIKADKKI----- 1339
XX
XX 177 PEIHHPNRRVEEEDSOOKEDESVPEPLIOHOMKDPDSQENYSDDEPTNAS-----VPT 229
XX DB 1340 -KIQINQATDEIEQENARK-----IEEAKIEA--KQNIQNSTRDOVNEAKTGINKIEMI 1393
XX
XX 230 PLHATTKPTFAQL--NKUNE-----VNSEPEALTKM----- 260
XX DB 1394 TPATTVKSEARQAVONKANEOQNIHQNPDPATNEEQEALINVSALARVQAQINAEHTT 1453
XX
XX 261 -----LKEENPSNLS-----LDEKVALYLSPTNNNSKAVSDMDSHL--ON 299
XX DB 1454 QGVKTIKIDATISLRINAQVEKESARNAIEOKATQOTQFINND--NATDEKEEVANN 1511
XX
XX 300 LODASKNKTNTENIHLSFALKAPKNDIENP-----LNSLTNADISLRSSSSSSSLSQSLN 355
XX DB 1512 LVIAIKQKSLDINLSL-----SNDDEVAKAVAGINEIAN----- 1546
XX
XX 356 DNRVLESVGPSPKAVNPGSLNDGIGKFSDEVESLLPRDLSRDKLETTKEHDAP----- 410
XX DB 1547 -----VLPTAVVSKAKKIDQKLAQOI--MQIQTHQATTEKEEAALQOLAN 1591
XX
XX 411 -----EHNNEFIKAKSTNTNKQGLV-----SSDHLDSFDRSYNHTQESILN 454
XX DB 1592 OKSNEARTALONEHNSNGVAQAKSNGIHEIELVMPDAHKKSDAKOSIDMKYN--EOS--N 1647

QY 455 LNSA-----SOSQISLNALE-----KQRTQDEQEQAAPEE-ETSESDNIKVKQ 500
DB 1648 TITPTDPADEEKQKALDKLAKDAGYNNKVDDAQNTQVSDAKTEKIDITITTIQANVAK 1707
QY 501 EPKSNLEF--VKVTKKEPVSAATEIKAPKREFSSRIIRIKNEDELAEPADHPKNEANS 559
DB 1708 KPSARVELDSKPEDLKRQINATP--NATEEKOALQRLANKRQEVNKLINQDRDNEVEQ 1766
QY 560 H---VEDTDALKKALNDEESDTTQNSTKMSIRPHIDSMDKLEDSNDGREDNDISR 615
DB 1767 HKNIGQLELETI--HANPFRKSDALO--ELQTFE-ISQTELLNNKADNEKEDEAKR 1819
QY 616 ---FEKSDILNDV--SOTSDIIGDKYGNSSSEIT-----KTL 648
DB 1820 LLEISKNNKTTTNNQATQNNQVDAKONGNEIATITPATITKTDAKTAIDKKAQOVTI 1879
QY 649 APPRSNDNDKENS-----LEDPANNESILOOLEVPHTEKEDDSILANSSNIAPPEE 700
DB 1880 INGNNDATDEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 1937
QY 701 LPLVVEANDYSSFN-----VTKTDAVSSFE-----ESLSREHET 737
DB 1938 -----TKTNAKQKINDKAQOLIOINNTPTATEEKEQEAATRVNAGLAQIOMINNAHST 1992
QY 738 D-----SKPINFISIMHKQEKQ-----KHQHKVP 763
DB 1993 QEVNESKTSIATIKSVQPNVIRKPPAINSLTQDANNQKLTIGNDGNATDDEKEAKOLV 2052
QY 764 TKOIIASVYO--YKNEQESRVTSDK-----VKIPNAIQFKFEVYVMSRRVSPMDDL 816
DB 2053 TOKLNOIQIKIHESITODNOVDNKAQAITAIKLINANAKROAINILT--NLESKSDI 2111
QY 817 NVSQFLPELSEDSGFKDLN--FANYSNNTNRPSPFTPLSTKYNVLSINDPNVPEPEPK 874
DB 2112 RAMQDPTTEKKNPAIOSIDTTLQARNNINGANT--NALVDENLEDGQKQRLQIVLSTQTK 2170
QY 875 SYAEINARLSANKAPNAPPLPQROPSSRSRNSKRVSRFRVPTFIR--RTSALA 933
DB 2171 TOKKADIQAQIGQKSTIDP-----NONATTEKQKLELNOFTNGVNDRIQALAA 2222
QY 934 ---PCDMYNDIEPDDFGAGSKPTI-----RAEGKATLPMSDKD--DVKRIINA 975
DB 2223 NQNVTEKKNILLETI--RNVETPIYVPRKANEIIRKKAABQTTLLNQNDATLEKQIALG 2281
QY 976 KKGVTODEYIN-----AKLVQDKP-----KNSIVTDPEDRYEE 1009
DB 2282 KLEEVKNEALNOVSOAHSNNDVYKIAENNGIAKISEVHPETIIRKNAKOEIEQDAQSOIDT 2341
QY 1010 L-----QOTASIHATIDSSIYGRPDSI-----TDLPLYLSD 1043
DB 2342 INANKSTNEEKSAALIDRVNAVAKIDAINNTNATTTQOLVADAKNSGNTSISQILP--STA 2399
QY 1044 LKKPPTALLSADRLFMEQEVNPLRSNSVLVH--PGAGA-----ATN--SSMLPEPEDELIN 1095
DB 2400 VKTNALALASE-----AKNKNAIIDQTFNATAEKEEKANNNKVDRLQEEADANIILK 2450
QY 1096 SPARANSNSDNVAISG--NASTISF-----NOLDMNPDQATIGKIOEQPASKSA-- 1145
DB 2451 AHTTDEVNNTKNOQAVONINAVQEVIRKQNVKQNLQNFIDNQKIIENPDPATLEEKAA 2510
QY 1146 -----NTVGGDDGGLASAPETPTPKKESISSKPAKLSSASPRKPIKIGSVYRIKKN 1200
DB 2511 NRILOVNLVISTSEIANVDHN-----NEVDALDKARP-----KTEETIYPOVSKK 2555
QY 1201 GSIAIGIEPIPKATHKPKKSFQGNIEISNHKVRDGISPSSESEHQHNPWSVPSQYTD 1260
DB 2556 RDV--LNAIQEAFNSQTOIEQENQEAPENEKFTALN-----KINQOLLQKAVNIDQ----- 2604
QY 1261 TSTVPEKQV-----QHKPREKQKQKHHNHHNHHHKKQTDI--PGVYVDE 1305
DB 2605 ---AOSNNDVDSAKTRSIQDIEQIOPHPQTKATGTRHLEKNAQOOSTATATPNSSTIEE 2660

```

QY 1306 IPDVG--LOE-----RGR-----LEFRLGJIKNINLP-----DINTH 1335
Db 2661 RQESATLQEVLLKALKALAKIDKGTNDVEKTYVNGIAETENI-LEATTYKOKAKADVNAE 2719
QY 1336 KGRFTLLDNGVHCYVTPREYNMDHNVAIGKEFELTVADSLFEIL-"TLKASYEKPRCTL 1393
Db 2720 KEQNINQINSNDEATT-----BEKLVASONLNHVYETTNQMIEDAPRTNQ 2764
QY 1394 VETYEKK-----VYKSNRLSLRFGS-----KDITTTKEVPTEYKDTWANKFA 1437
Db 2765 VNEKKNNGIGTIRDIQPLVYKKPTPAKSKIESAVEKKETJENOTONATHDEVRE-GLNQLN 2833
QY 1438 PDGSEFARCYIDLOQFEDQITGKASQSPDLNCFN-----EMETMSNGNGMPKRRCKPKYIAOLE 1493
Db 2824 QIHEKAKNDVNOQSOTMOOYE-NEONSLSDOJNNFRPDESKKNNAVAELTVKAOQNKIDETE 2882
QY 1494 VKM--LYVPRSDREILPLPSIRASYESIMELN--NEONNYFEGCYLHOEGGDCPIFFKRKF 1549
Db 2883 QEFSATQEEKDNMLQHLNDEBYKKEIINSINQANTDNEVDNAKTSGLNITTEVRPEYKMK-- 2940
QY 1550 KLMGTSLLAHSEISHTRAKINLSKYVDLLVYDKENIDRNSHNRNESDYLLLDHAFKIKFA 1609
Db 2941 ---KNALIKLYVDSPTQEAIIIN-----GYPDATEDEQLQEAANSKLKILDDAKKQIGLA 2990
QY 1610 --NGELIDFCARPKHEMK 1625
Db 2991 HTNNEVDYDLYNEVSQMK 3008

```

	RESULT 9
AAAG82935	ID
AAAG82935	standard; Protein: 5024 AA.
XX	AAAG82935:
AC	
XX	03-SEP-2001 (first entry)
DT	
XX	S. epidermidis open reading frame protein sequence SEQ ID NO:2964.
DE	
XX	Staphylococcus epidermidis SMI strain; infection; diagnosis;
KM	vaccination; endocarditis.
WM	
XX	Staphylococcus epidermidis.
OS	
XX	MO200134809-A2.
PN	
XX	17-MAY-2001.
PD	
XX	09-NOV-2000; 2000MO-US30782.
PF	
XX	09-NOV-1999; 99US-0164258.
PA	
XX	(GLAX) GLAXO GROUP LTD.
PI	
XX	Kimmerly WJ;
DR	
XX	WPI; 2001-316485/33.
DR	
XX	N-PDB: AAH53785.
PT	
XX	Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
CC	useful for vaccinating against infections, e.g. endocarditis -
PS	Claim 18; Page 779-781; 218bp; English.
XX	
XX	AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC	(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC	(I) and (II) can have antibacterial activity and therefore can be used
CC	in vaccination. The nucleic acids (I) may be used to produce the
CC	S. epidermidis polypeptides (II) via the production of vectors
CC	containing them which are used to produce hosts cells which express the
CC	polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC	used to vaccinate subjects and to raise antibodies against the bacteria.
CC	The polypeptides may also be used to assay for other inhibitors of their

CC treatment and therefore identify compounds that may be used for the
CC treatment of *S. epidermidis* infections, e.g. *epidermiditis*. AAH53971 to
CC AAH53980 represent specifically claimed *S. epidermidis* genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472
CC no sequences are present for SEQ ID NO:4455 to 4464.
CX

Query Match	3.9%	Score 338	DB 22	Length 5024
Best Local Similarity	19.7%	Pred. No. 3.8e-11		
Matches 388	Conservative 309	Mismatches 714	Indels 560	Gaps 91

[illegible]


```

Db 2917 EEEQKHHPKSSKSSISSEQOPSTEQEIVSYTEHDLKP-----EEKPPYVOYIOSET 2972
Qy 248 EV-----NSEPPA-LITPMKLRNFMSL 270
Db 2973 NVEETKDDTGVKHQVYTKRMLRRPAGGEIETIEVARDPOPEAEITIVETPE---FVN 3029
Qy 271 LDEKVNLYLSPFTNNNSKNVSDMSHLQNLQDASKNKTN-ENHNLSEFALAPKNDIENP 329
Db 3030 QDEKPK---EFKKKTRKRYKKDDIDHYIOKLTIELETPKTELEKYEKEIEPEIVKOKPLDSP 3086
Qy 330 LNSLTNADISLRSSGSSOSSLSLRND-----NRVLESYPGSP----- 367
Db 3087 IDVLDESPEKQVKOKKSRSTKVPNEETPVQEQYAKVNVVEEAPQEPEIYQILEVYKV 3146
Qy 368 -----KVNPGSL-----NDGKGFSEVESESL 392
Db 3147 EVDKVEYTEDGKRVQEKTKRYLAKTIGPEQOTFKITMIESEDNDSTVYVDEPELIAS 3206
Qy 393 PROLSRDLKLETKKHDAPENHNENFIDAKSTYNNKGLVSSDDHLSFDRSVNHTQESI 452
Db 3207 POSI-----EHH--PEQSEKELAPKPKTVRK---VKDD-----L 3237
Qy 453 LNLNSASQSOISLNALEKQKQOTQEQTOAAREEETSFSDNKKVQEKSNLEFVKVT 512
Db 3238 SDYVKKLIEETIPKVDLEKYEVE-----MPEKPKLVSDSIEPEKPD----- 3282
Qy 513 IKKEPVGA-TEIKAPKREFSSRLIRKNE-----EIAEPAD 548
Db 3283 -KSQPIVLDPDTKPKTKTPKTEDETDQVDPDEPTTVDTTDIPELPQTQAOED 3341
Qy 549 -----HPKKENANSIVEDTALLKALANDESDTTONSTKSIHHDSDMKLEDN 603
Db 3342 TATQAITPSAQEESSTQ-DQTKDTIOKTVKHKTKRPTOKSVETSELPVHKDQOISIH 3400
Qy 604 DG--DREDNDQISFEKSDILNDVSOQSDI---IGDKYGNSSSETTKTALPSPDNNDK 658
Db 3401 EELVEEQPEKILEVRIYDEVAEYESQPIVEEVEDEPPAIEYEDVYKPKS----- 3455
Qy 659 ENSKSLDPAANNESLQOOLEVPHTKEDDSILANSSNIAPEELTPVVEANDVSSFNVDY 718
Db 3456 -----KKKVVKKKTDH-----DELKKMLE-----GEIE 3481
Qy 719 KT-FDAVSSFEESISREHEDSKRINFISTHKOEOKKQIHKVPYKQIATISAOQKNE 777
Db 3482 KTELEKYEKIEFDVPKLKEPFALEPIKIERKQK-----PKVTL----- 3524
Qy 778 QESRVTSDKVIAPAIQFKKFK-----EYNV-----MSRNVSPMODLVNSQFLPE 824
Db 3525 -----DATDVPTKTVALKSKRKEKPAEELTVQLPKFRLKARV---LVEYPPAPLIPK 3574
Qy 825 LSEDSGFKDLNFANYSNNTNRP---SFTPLSTKNVLSNDNDPNVVEPPEPKSYAEIRN 881
Db 3575 TTDTGAIKD--NGELSHNIEAEELIKFKPKHTKKI--KKIDOLEKVELEKYEK----- 3626
Qy 882 ARRLSANKAPNQPAPLPQROPSTSRNSNKKRYSRRVFFELIRRSSALAPCDMINDI 941
Db 3627 ---ISSEEPPEKTPYKKPEKAPK-----PEEKEDV 3655
Qy 942 FDDPGAGSKPTIKAEG-----MKTLPMSDKDDVRIINAKKGVTODEYINAKLVQOKPK 996
Db 3656 KIKGKGGKKKPEEAEENVTLKNIPQKPO-EVEEELVKOKPREVELEVQO--TKRPKD 3712
Qy 997 NSIYTPDEPDEYELQQTASIHMATIDSSIYGRPDISTDM---LPLYSDELKRP-PTAL 1051
Db 3713 GEFVVEP-----FEPSEFDRPREYVDELQIHEPKEPKKPSKTKY 3755
Qy 1052 LSNADRLMEQGVHPLRSNVLYVHPGAGATNSSLPPDEPFLISPARANSNNDWVAIS 1111
Db 3756 KPKDKSKSEPE-----TIVSELVAGVPPKEEAPIPQDVK-FRKPDEDAEEEDT----- 3802
Qy 1112 GNASTISFNOL-----DMNFDQATIGOKIOEOPASKSANTVRDDDGSLASAPTPPT 1166

```

```

Db 3803 ---SEIKLRVPQASKDENPDEQALVTPK-AEERIPQETEDKALIDKKPKSKKKVQ 3858
Qy 1167 KKESSISKRAKISSAPRK-----SPIKIGSP--VRVYIKNGSIGAGIEPIPKATNPKPK 1218
Db 3859 KEQELAKEPEPEFEVSVKEEELVNDKPIEIEKPKDVKKKEKKPKAPVSEVNVLEEKPK 3918
Qy 1219 SFQGNELIS-NHKVNRDGGISPSGSEHOQNPMSVSPSQYDATSTVDEKNDVOHKPRE 1277
Db 3919 EEPREIPEVEYKITTVLEPDAF--KEHQVKVIDFDRQETTEEVI--EEKVYTRKKKP 3974
Qy 1278 KOKQKHHNHHHHHKKQTDPIPGVDDDEIPVQLOERKLEFRVLGGINILPDIINTHKG 1337
Db 3975 KPQPEEEFVTLKEPKEQIOPDVVASAI-SLPIEE-----PEQKPYQ 4017
Qy 1338 RFTLLDNGVHCVTTPREXMDHNVAIKGEFELVADSL--EFITLKASYEKP--RGTL 1393
Db 4018 EVELKITQ-----TTPF-EPNDVQIAYEAKTKRYPKKVYKEDKIYVVAEESKORVEETI 4071
Qy 1394 VEYTERKKVYSRNLRLSRLGSKDIIITTKFY--PTEVKDTWANKFAPDGSFARCYIDLQ 1450
Db 4072 VEV-EKOEKKKSEKPKSYEER--ISETQISEKPIEVAEE-----APEET-----PK 4116
Qy 1451 QFEDQITGK--ASQFDLNCFMEMETMSNGQPMK-----RGKPYKIAQLVYKMLYVP 1500
Db 4117 VEEKVAKKEPDSYEETLKEETDEKVTIVDQPEEAPVAVVKKRKEPAVEAEFV--- 4173
Qy 1501 RSDPREILPTISRSA 1515
Db 4174 MTEPRIVETSVEYA 4188

RESULT 11
ABB60327
ID ABB60327 standard; Protein; 3111 AA.
XX
AC ABB60327;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7773.
XX
KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN MO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW.
XX
WP1: 2001-656860/75.
XX
DR N-PSDB: ABL04430.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 7773; 21pp + Sequence listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

```


XX (ELITRA PHARM INC.
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
PI Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
DR N-PSDB: AAS55262.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
PS Example 3; Seq ID No 12996; 511bp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6281 AA:
SQ
Query Match 3.8%; Score 327; DB 22; Length 6281;
Best Local Similarity 17.7%; Pred. No. 2.5e-10;
Matches 296; Conservative 279; Mismatches 607; Indels 488; Gaps 71;
42 NSKPSIDPSSSDPTYSQ--DOEGK-----KEKKD-----71
Db 2459 NSGPNTTAAADVTATGQRNNAETGLNGDTNLATAKQKQKALRQMTLSDAKOSITG 2518
QY 72 -----TAFQTSFDRNFDLNSID-IDQOTIQHQOQOPOQOQOOLSOTDNLIDFSFOT 122
Db 2519 QIDSATQVGVGQSVKDNATWLDNAMQLRNSINKDKDVAKSPYV--DADRRKQNAUNT 2575
QY 123 PMNST---IDLTKONPTVQVKNENHAPTYLNTSPKNSIMKATPKASPKKATVTPPEI 179
Db 2576 AVTMAENIINATISQ--PTLDSAVTQAAANOVST--NKTALGAQONLAN--KKOETTANINOL 2631
QY 180 HHYRDNREVED-----OSQKEDSVEPRLIOHWKD--PSOFNYS 218
Db 2632 SHL--NNAKQKODLTOVTANPNISTVQVYKTAEOLOQAMERLINGLODKDYKOSVNT 2689
QY 219 DEDTNAVPRPTPLHTTKPTFAQ--LINKNEVNSEPE---ALTDKLRKRENSNLSLDE 273
Db 2690 DAD-----DEKQTAUNNAVTAAENIINQANGTANOSQOEAALSTVTTTKO---ALNGDR 2741
QY 274 KYVLYLSTPTNNNSKVVSDMSHLQNLQDASKKKTNTNHNLSFALKARNDLENPLNSL 333
Db 2742 KY-----TDKNNANQTLSTLD-----NLNNAKQAGVATGNI-----NOAHVY 2778
QY 334 TNAIDSLRSGSSSSLOSLRDNRVLESVPGSPKKVNPGLSLND-----GIGFSFDEVY 388
Db 2779 AEVTAQIOTQAOELNTAMGNLKN-----SLNDKOTTLGSONFADA-- 2817
QY 389 ESLLPRDLRSKLETTKEHDARENNENFIDAKSTNTN--KGOL-----LVSS 434
Db 2818 -----DPEKKNAVNEAVNANENILN--KSTGTNVPKQDVEAAMQVNAVNTKALNG 2865
QY 435 DHHLSDFDSYNTHTEOSIINLNSASQS-----QISLNALEKOROTQOEOQTOAEPREE 489

Db 2866 TQMLEKAKQAHNTAIDGLSHLTNAQKEALKQOLVOOSTTVAEAGNBOKANNDAAADKLR 2925
QY 490 TSPSDNITKQOEKPSNLEPVKVTIRKEPVSAITEIKAPKREFSRILIRIN--EDEINPAD 548
Db 2926 QSIADNATTKQ-----NONYTDASQNK-----KAVYNAVTTAAGIIDDITTSPT- 2969
QY 549 IHPKRENEANSHEVDLALKKALNDESDPTQNSTKMSIRFHIDSKWLEDSNGDRE 608
Db 2970 LDPTVINQAAQGVSTT-----KNALNGENLEAKQOASQSL-----GSLDLNNAQKQ 3018
QY 609 D--NDDISREKSDIINDVQOTSIDIIDKYGNSSETTTTL-----APPRSN 655
Db 3019 TVTDQINGAHVDEANQIKQMAQNLNTAMGNLQALADDAFKATVNTFDADQAKQAAVN 3078
QY 656 NDKENKSLDEPAN-----NESIQOOLEVPHTEKEDSILANS-- 693
Db 3079 TATVNAENIISKANGNATQAEVEQAIKQVNAKQALNNGNANVQHAKKDEATALINSNDL 3138
QY 694 NIAPREELTPVEANDYSFNDVTKTF---DAYSFEBSLRHEHTDSKPINFISTWH 749
Db 3139 NQAKQDALQOQVONATTVAGVNVKQTAQELNNAQTOLQGIADKQOT----- 3186
QY 750 KQKQKQHQIHKKVPTKOIIASVQYKNEQESRVTSPKVIIPNAIQKFKREVNVMSRVV 809
Db 3187 ---KADGNFVNADPRQMAVNAKAEALISATPVDVVTPEI-----TALNKV 3234
QY 810 SPDMDDLNVQSFLEPSEDSGFKDLNFANVSNNTNPRSFPTPLST--KNVLSINDNDPV 867
Db 3235 TQAKNDLN-----GNTN-----LATKQVNVQAHIDQPL 3264
QY 868 VERPEPKSYAETRNARLSANKAPQAPPLPRQOPSSSTRKSNKRVSRFVTEIRR 927
Db 3265 NQAKRDEYSKQITQV-----TLVNVVNAIQQAATTLNDAMTOLQGIANKKQIIG 3314
QY 928 TSSALAPCQWYNDI--FDDFGAGSKPTIKAGM---KTLPSMDQDVKRLIAKKGVTOE 983
Db 3315 SEN-----YHADPTKQATVNAVTAKAEBLLKQTTNPTMDNTTQOALTQKYNQNL 3367
QY 984 YINAKLVQK-----PKNSIVTPDRRELELOOTASINATIDSSISYGR 1028
Db 3368 NGNQKLADAKQDAKQKTLGLTLHLDQAQKALT---QVQAPADIATVNVNQNAQNL-- 3421
QY 1029 PDSISTDMLRYLSDCLKKPTALLSADRLMEQEVNRLSNVSLVIRGAGATNSSMLPE 1088
Db 3422 ---NNAMTNLNNALODKTETLNSINFDDADKADATVNAHSAEGILSKANGSMAQO 3476
QY 1089 PDFF-----LINSFARNVNSNDVVALSGNAS--TISFNQDMNDDOATIGQIOEQPASK 1143
Db 3477 TEVEQAMQRYNE--AKQALGNDVQVAKDAKQVITNANDLNOQDALKQVDAQVYA 3535
QY 1144 SANTVKGDDGLASAREPTPTPK--KESISSPKAKLSASPRKSPRTIGSPVVRVIRKNS 1202
Db 3536 NVNTIK-----QTADLNOAMTQLOKGIADKQO-----TKANCN 3569
QY 1203 IAGIERIPKATIRPKKSPFGQNEISNKKVUDGGISPSGSEHQOANPSWVSQYTDATS 1262
Db 3570 FVNAD-----TDKQNAVNAVA-----HACQII-- 3593
QY 1263 TVPDEKQDVQNKREKQK--QKNNHNNNNNNKQKTDIGVUDELIDVGLDGRGKLYFR 1320
Db 3594 GTPNAVNDVQVAAQADQVNOQAKGLNGMNLQAVKONANTADQLP----- 3640
QY 1321 VLGIKINILRDIPTNKGREF--LTLDGVCVUTPREXNMDDHVAIKREFELVADSLFT 1379
Db 3641 ---NLNORQKALKDOVSHAEVLTGVNAI---KQMAADLNNAMG----- 3678
QY 1380 LTLK---ASYEKRGTLVEVTEKKVVKSR-----NRLSRLGSKDITTTTKVPTPEVK 1429
Db 3679 -TLKQOIQANNSQVPS--VDFEQADODKQOAVNNANQAOQIANG-----IPTPV- 3725
QY 1430 DTWANKFARDGSPARCYTIDLOQFEQDITG--KASQDILACFNEMETMSNGNQPMKKGKPY 1487

QY 378 DGKGFSEV---VESLP-----RDLSDKLETTK----- 405
Db 859 VDAKGIQNLIGHLTAKEKVVPOTKLSPSTKEASKALPRKLVTRPVLKSRSTKVGOK 918
QY 406 -EHDAPR-----NNEFIDAKSTNTKQGLVSSDH----- 437
Db 919 NPSEKPVHEEFLHPVSAVAGNRIPTSSISDDGSIPLGFDNEGKRSKTPVRSQRYKOK 978
QY 438 -----LDSFDRSYNHTEOSILN-----LNSASOSQISLNALEKQROTQEQTOAE 485
Db 979 INETDISDDATPDYDETEEEQPSIEIQPVKNSAADSKTFPRHQVYGAFFSDSDSDAG 1038
QY 486 PEETSFSQDNKVKOEPSNLE---FKVYTIKEPVSAIEAPKREFS-----SRI--- 534
Db 1039 PSSPMDFGDMNSIEDSRTSEEDVTAMAKLVNENEKPFVDEVLSPSKSGTEATVDCSIDD 1098
QY 535 -LRINDEDEIAPADIHPKKEANEANSHVEDTDLKALNDEDESTTONSTKMSIRFI 593
Db 1099 PLKTGGEATSSHSEITPCASSEKCVTDE---KQPIKKEKSDITLDRKSEKSRMST 1154
QY 594 DSDMKLESDNDGRDNDISRFKESDILNDVSDIIGDKYGNSSSEITTKTLAPRS 653
Db 1155 KSLMTTDEH-----VKDSSTTSKI-----NEKRLTSSPLKTPGS 1189
QY 654 DNN-----DKENSKSLEDPA---NNESLQOOLEV----- 679
Db 1190 SENTELRLLEGNASDKQNEHLISPEALTNQNAKLKNELEVSPKPYSLRSRGRKS 1249
QY 680 -PHTKEDDSILANSSNIA---PPEELTLPVANDYSSFNQDYKTFDASSFEEISLREH 735
Db 1250 GSHNKKLDESEKHSNISCAPKEEKATERLEDGEOULTYVDSTE----- 1293
QY 736 ETDG--KPINFISIMHKOQKQKHOIHKVPYKQIIA--SYQYKNEQESRYTSKVKIPN 791
Db 1294 DTDSTRNPIDNTAENTAIIRAKR-----KTRCIYKSKOSKEMKDLKEDGAVSSKDDL- 1346
QY 792 AIOFKKFEVNNMSRRVVSPPMDLNVSOFLPELSEDSGFKDLNFANYSNNTNRPRTPT 851
Db 1347 -----PTMDS-----SEKPEDOKLDVDPVSVKSGTR----- 1374
QY 852 LSTKAVLSINDPNNVPEPEPKSYAEIRNARLSANKAPNOADPLPQROPPSSTRSNS 911
Db 1375 ---KSLRASLNNEVPAIQ-PEP-----VINTVSSSENSTSKME---DYVTEPNDTK--- 1418
QY 912 NKRVSRFRVPTPE---IRRTSSALAPCDMYNDIFDDFGAGSKPTIKAGMKTLPMDKD 967
Db 1419 NKRHTRNKSPQIKGSPPTIKHFESIESEAIKGSVDSSKAESSIRPTRQORSY---SKAKS 1475
QY 968 DVKRLNKKKGYTODEYINAKLVD-----QPKKNSIYTDPEDRYEEIQQTASIH 1018
Db 1476 EPKALADSKSENVSDPSISTYTQDLGQAKYDSEKPSGDVQNLNPKYDSCNE--QDACKDSN 1534
QY 1019 ATIDSSITGRPDSISTDMPLPYLDELKRPPTALLSADRLFMEOVHPLRSNSVLVHPGAG 1078
Db 1535 QAKD-----LPSKRRKROSQSGSTKIYPPVSQPKPEKLHEONRDESVRA----- 1579
QY 1079 AATNSSMLPEPPELINSFARNVSNNSDVAISGNASTISFNQLDMNFDDQATIGKIOE 1138
Db 1580 -----PKBOLDPTTSKNSRKGKRSIATNTPTMLTASE---DPEAVEGDSV-E 1627
QY 1139 QPASKSANTVRCDDGLASAPETPR-TPTKKSISKPAKLSASAPRK-----SPKITG 1191
Db 1628 VPIKRRKASTRRG-----SVPRDTPISSYVNASDESAVLESV--PKGGRSNISSVPAD 1677
QY 1192 SPVRVIKKNKSGIAGIEPIPKATHKPK 1217
Db 1678 TTIPKSSKSDSLNAIEIIPICEPR 1703